

Thu May 8 16:14:00 2003

us-09-658-621b-2.rat

Page 1

GenCore version 5.1.4_p5 4578
Copyright (c) 1993 - 2003 Comugen Ltd.

CM protein - protein search, using bw model

Run on: May 1, 2003, 23:09:26 ; Search time 25.5357 Seconds
(Without alignments)
570.352 Million cell updates/sec

Title: US-09-658-621b-2

Sequence: 1 MPTGTOSPFLLTLVLTIV.....NGSSLSYTPAVATSNL 495

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: Issued Patents AA: *
2: /cgn2_6/prodata/1/aa/5A.COMB.pap: *
3: /cgn2_6/prodata/1/aa/5B.COMB.pap: *
4: /cgn2_6/prodata/1/aa/6A.COMB.pap: *
5: /cgn2_6/prodata/1/aa/6B.COMB.pap: *
6: /cgn2_6/prodata/1/aa/backfiles1.pap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	2019	78.8	2035	2	US-08-479-537A-2
2	2019	78.8	2035	4	US-09-083-116-2
3	2019	78.8	2035	4	US-09-134-916A-2
4	1091	42.6	1867	2	US-08-479-537A-5
5	1091	42.6	1867	4	US-09-083-116-5
6	234.5	9.2	907	4	US-09-134-916A-5
7	234.5	9.2	907	3	US-08-783-774-2
8	234.5	9.2	907	4	US-09-328-559A-1
9	234.5	9.2	907	5	PCT-US95-04611A-19
10	228	8.9	878	4	US-09-526-706B-2
11	207.5	8.1	750	4	US-08-063-229A-4
12	205	8.0	40	2	US-08-288-059-1
13	205	8.0	1060	4	US-08-911-193-2
14	198	7.7	2137	4	US-09-134-001C-4463
15	198	7.7	941	4	US-07-757-022B-14
16	192	7.5	1022	4	US-07-757-022B-84
17	192	7.5	1038	4	US-07-757-022B-74
18	192	7.5	1049	4	US-07-757-022B-56
19	192	7.5	1140	4	US-07-757-022B-104
20	192	7.5	1311	4	US-07-757-022B-44
21	192	7.5	1311	4	US-07-757-022B-42
22	192	7.5	1311	4	US-07-757-022B-142
23	192	7.5	1314	4	US-07-757-022B-50
24	192	7.5	1320	4	US-07-757-022B-46
25	192	7.5	1320	4	US-07-757-022B-60
26	192	7.5	1354	4	US-07-757-022B-48
27	192	7.5	1354	4	US-07-757-022B-48

28	192	7.5	1361	4	US-07-757-022B-40
29	192	7.5	1363	4	US-07-757-022B-52
30	192	7.5	1404	4	US-07-757-022B-2
31	192	7.5	1404	4	US-07-757-022B-62
32	189	7.4	805	4	US-09-103-429A-4
33	189.5	7.4	786	4	US-09-513-429A-3
34	174	6.8	1125	4	US-09-513-783A-152
35	174	6.8	1185	4	US-09-041-866-23
36	167.5	6.5	1610	4	US-09-513-783A-22
37	167.5	6.5	2972	4	US-09-579-181-2
38	167.5	6.5	3118	4	US-09-579-181-1
39	165	6.4	401	6	5252556-1
40	162.5	6.3	1581	4	US-09-110-517-2
41	162	6.3	267	4	US-08-818-112-142
42	162	6.3	267	4	US-08-818-111-137
43	162	6.3	267	4	US-09-056-556-142
44	162	6.3	267	4	US-09-072-596-137
45	161.5	6.3	1719	2	US-08-459-568-4

ALIGNMENTS

RESULT 1
US-08-479-537A-2
Sequence 2, Appl 1
Patent No. 5813181
GENERAL INFORMATION:
APPLICANT: CHAMRON, Pierre
APPLICANT: KIRBY, Marie-Paule
APPLICANT: LARHE, Richard
APPLICANT: HAREUYENT, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DONNE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
City: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 2213-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids

1	TYPE:	amino acid
2	STRANDEDNESS:	single
3	TOPOLOGY:	linear
4	MOLECULE TYPE:	peptide
5	FEATURES:	
6	NAME/KEY:	Peptide
7	LOCATION:	128..1899
8	OTHER INFORMATION:	/note= "The amino acids spanning
9	OTHER INFORMATION:	128 to 1899 constitute a repeated region wherein the repeat i
10	OTHER INFORMATION:	20 amino acid, 17 of which are fixed. The number of such
11	OTHER INFORMATION:	repeats varies from 1 to 40."
12	NAME/KEY:	Peptide
13	LOCATION:	134
14	OTHER INFORMATION:	/note= "Amino acid 134 is X1 = Xaa
15	OTHER INFORMATION:	Xaa,Xaa which is the codon for Pro or Ala wherein Pro = CCT,
16	OTHER INFORMATION:	CCC, CCA, and Ala = GCT, GCC, GCA, or GCG."
17	FEATURES:	
18	NAME/KEY:	Peptide
19	LOCATION:	144
20	OTHER INFORMATION:	/note= "Amino acid 144 is Y = Xaa
21	OTHER INFORMATION:	which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG
22	OTHER INFORMATION:	or ACG; and Asn = AAT or AAC."
23	FEATURES:	
24	NAME/KEY:	Peptide
25	LOCATION:	147
26	OTHER INFORMATION:	/note= "Amino acid 147 is X2 = Xaa
27	OTHER INFORMATION:	which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCM
28	OTHER INFORMATION:	or CCG; and Ala = GCT, GCC, GCA, or GCG."
29	FEATURES:	
30	NAME/KEY:	Peptide
31	LOCATION:	12
32	OTHER INFORMATION:	/note= "Amino acids 1 to 21 are a
33	OTHER INFORMATION:	21 amino acid precursor sequence."
34	OTHER INFORMATION:	
35	US-08-479-537A-2	
36	Query Match	78.8%; Score 2019; DB 2; Length 2035;
37	Best Local Similarity	85.3%; Pred. No. 9,7e-146;
38	Matches 399; Conservative	5; Mismatches 62; Indels 2; Gaps 1;
39		
40	QY	28 SSTPGGEKETSATORSGSVSTENNAVSMVTSSVLSSHPSGSGSTTGOODVTLAPATEPA 87
41	DQ	1570 STAFPAHGVTSADDKPKFGSTAFAHCVTSAPADKRPPGSTAPXHG--TTSAPDPRPX 1627
42	QY	88 SGSAATMGQDVTSVPVTRPALGSTTPPADVTSARDNKPAGPSTAPPAHGVTSAPDRPP 147
43	DQ	1628 PSTAPPAHGVTSAPDRPXPGSTAPPAHGVTSADDKRPKPGSTAFAHGVTSAPDRPX 1687
44	QY	148 POSTRAPAAHGVTSAPDRPARPGSTAPPAHGVTSAPDNKPFALGSTAPPVHNYSASGSAG 207
45	DQ	1688 POSTRAPAAHGVTSAPDRPARPGSTAPPAHGVTSADDKRPXLSTAPPAHNTYSASGSAG 1747
46	QY	208 SASTLPYNKGSAATTTPASKSTSPSIPESHSDVPTLASHTSKTDASGTHRSVPLEPLTS 267
47	DQ	1748 SASLTVINGMSAKATITPASKSTPPSLFESHMDIVPTILASHSKTDASGTHRSVPLEPLTS 1807
48	QY	268 SNHSRSTQLSTGVSFPFLSPFIHSLQNSLSLDBPETDYQLGRDISMPLOANKGGSL 1867
49	DQ	1808 SNHSRSTQLSTGVSFPFLSPFIHSLQNSLSLDBPETDYQLGRDISMPLOANKGGSL 1867
50	QY	328 GSINIKRRPPGSVVOLTLAPREGTINVHVEVFQFOXYTAARVLTISPVSVSHVPP 387
51	DQ	1868 GSINIKRRPPGSVVOLTLAPREGTINVHDETFQFOXYTAARVLTISPVSVSHVPP 1927
52	QY	388 FASGSGVGVMQIALVLVLCVLAALAIVYLIALAVCCCKRKNYGGLDIPARDTYHPMS 447
53	DQ	1928 FASGSGVGVMQIALVLVLCVLAALAIVYLIALAVCCCKRKRYGGLDIPARDTYHPMS 1967
54	QY	448 EYFTYHTHGRVYPSSSTDSPYEKVASNGGSGSLSYTNPAVAATSANTL 495
55	DQ	1988 EYFTYHTHGRVYPSSSTDSPYEKVASNGGSGSLSYTNPAVAATSANTL 2035

RESULT 2
 US-09-083-116-2
 Sequence 2, Application US/09083116
 Patent No. 6203795
 GENERAL INFORMATION:
 APPLICANT: CHAMBER, Pierre
 APPLICANT: KIENEY, Marie-Paule
 APPLICANT: LATHE, Richard
 APPLICANT: HAREVUENI, Mara
 TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
 NUMBER OF INVENTIONS: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 STREET: BURNS, DOANE, SNECKER & MATHIS, L.L.P.
 STREET: 1000 Park 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22113-1404
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/083,116
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/479,537
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/FR91/00835
 FILING DATE: 23-OCT-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/039,320
 FILING DATE: 04-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/403,576
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Teskin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-6620
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2035 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 128-1899
 OTHER INFORMATION: /note= "The amino acids spanning
 OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
 OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
 OTHER INFORMATION: repeats varies from 1 to 40."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
 OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
 OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
 NAME/KEY: Peptide
 LOCATION: 144
 OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, TGT,
 OTHER INFORMATION: or ACG; and Asn = AAT or AAC "

FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2
Query Match 78.8%; Score 2019; DB 4; Length 2035;
Best Local Similarity 85.3%; Pred. No. 9.7e-146;
Matches 399; Conservative 5; Mismatches 62; Indels 2; Gaps 1;
QY 28 SSTPGKETATQKRSVPSSTKKNVMTSSVLSHSPGSGSTTQGDVTLAPATERA 87
DB 1570 STAPXAHGVTSADKRPXGSTAPXAHGVTSADKRPXGSTAPXAHG--VTSAPDXRPX 1627
QY 88 SGAATWQGVTSVPTSPALGSTPPAHADVTSAPDNKPAFGSTAPXAHGVTSAPETRP 147
DB 1628 PSTAPXAHGVTSAPDXRPXGSTAPXAHGVTSADKRPXGSTAPXAHGVTSAPDXRPX 1687
QY 148 PSTAPXAHGVTSAPDXRPXGSTAPXAHGVTSADKRPXGSTAPXAHGVTSAPDXRPX 207
DB 1688 PSTAPXAHGVTSAPDXRPXGSTAPXAHGVTSADKRPXGSTAPXAHGVTSAPDXRPX 1747
QY 208 SASIVHNGTSABATTPRASKSTPSPISPHSDPTTLASHTKIDASTHSTVPEPLS 267
DB 1748 SASIVHNGTSABATTPRASKSTPSPISPHSDPTTLASHTKIDASTHSTVPEPLS 1807
QY 268 SNHSTSPQLSTGVSPFPLSFHISNLQPNGLSDPSTVYQELQROISMPFOIKKGGFL 327
DB 1808 SNHSTSPQLSTGVSPFPLSFHISNLQPNGLSDPSTVYQELQROISMPFOIKKGGFL 1867
QY 328 GLSNKKPPSGVVVQTLTAFREGTINVHDETOPQVYTAASRYVLTISVSVSHVPP 387
DB 1868 GLSNKKPPSGVVVQTLTAFREGTINVHDETOPQVYTAASRYVLTISVSVSHVPP 1927
QY 388 FSNQSGAGVPGMGIALVVCVLAVALVYLTALVCCCRKRVGQDLFPARDYHPMS 447
DB 1928 FSNQSGAGVPGMGIALVVCVLAVALVYLTALVCCCRKRVGQDLFPARDYHPMS 1987
QY 448 EYPTHTHGRVPPSSSTRSPYKVSAGNGSSLSYTNPAVAATSAHL 495
DB 1988 EYPTHTHGRVPPSSSTRSPYKVSAGNGSSLSYTNPAVAATSAHL 2035
RESULT 3
US-09-134-916A-2
Sequence 2, Application US/09134916A
GENERAL INFORMATION:
PATENT NO. 6328956
APPLICANT: KIMMEL, Pierre
APPLICANT: KIMMEL, Pierre
APPLICANT: LAMPE, Richard
APPLICANT: HARRIVANT, Marc
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,916A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teekin, Robin L. 5,030
REGISTRATION NUMBER: 017753-025
TELEPHONE: (703) 816-6620
TELEFAX: (703) 816-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, AC
OTHER INFORMATION: or ACG, and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CC
OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-134-916A-2
Query Match 78.8%; Score 2019; DB 4; Length 2035;
Best Local Similarity 85.3%; Pred. No. 9.7e-146;
Matches 399; Conservative 5; Mismatches 62; Indels 2; Gaps 1;
QY 28 SSTPGKETATQKRSVPSSTKKNVMTSSVLSHSPGSGSTTQGDVTLAPATERA 87
DB 1570 STAPXAHGVTSADKRPXGSTAPXAHGVTSADKRPXGSTAPXAHG--VTSAPDXRPX 1627

Thu May 8 16:14:00 2003

us-09-658-621b-2.ra1

Page 4

QY 88 SGSAAWGDVTSVETRPALGTTTPEPAHDVTSAPDNKRPAGSTAPPAHGVTSAPDTRPP 147
Db 1628 PGTAPPAHGVTSAPDTRPPAGSTAPPAHGVTSAPDNKRPAGSTAPPAHGVTSAPDTRPP 1687
QY 148 PGTAPPAHGVTSAPDTRPPAGSTAPPAHGVTSAPDNKRPAGSTAPPAHGVTSAPDTRPP 207
Db 1688 PGTAPPAHGVTSAPDTRPPAGSTAPPAHGVTSAPDNKRPAGSTAPPAHGVTSAPDTRPP 1747
QY 208 SASLTVANGTSARATTPPAKSTPPSIPSHSDPTTLASHTKTDASSTHSTVPLTS 267
Db 1748 SASLTVANGTSARATTPPAKSTPPSIPSHSDPTTLASHTKTDASSTHSTVPLTS 1807
QY 268 SNHSTPQSTGVSEFFLSPHISNLOPNSSLEPSTDYQELORDISEMPLQYKGGFL 327
Db 1808 SNHSTPQSTGVSEFFLSPHISNLOPNSSLEPSTDYQELORDISEMPLQYKGGFL 1867
QY 328 GLSNIKRFGSVVQVLTAFREGTINVDVETCPNOYKTEASRYNLTISDVSASHVFP 387
Db 1868 GLSNIKRFGSVVQVLTAFREGTINVDVETCPNOYKTEASRYNLTISDVSASHVFP 1927
QY 388 FSAGSGAGVPGWGLALVLCVVALAIVYLIALAVCCCRKNGGOLDIFPARDTYHMS 447
Db 1928 FSAGSGAGVPGWGLALVLCVVALAIVYLIALAVCCCRKNGGOLDIFPARDTYHMS 1987
QY 448 EYPTVHTGRVYVPPSTDRSPYEKVASNGGSLSTYNPAVAATSNL 495
Db 1988 EYPTVHTGRVYVPPSTDRSPYEKVASNGGSLSTYNPAVAATSNL 2035

RESULT 4
US-08-479-537A-5
Sequence 5, Application US/08479537A
Patent No. 5621381
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREIVENT, Maira
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OF SEVERE TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 134-1867
OTHER INFORMATION: /note= "The amino acids spanning 134-1867 constitute a repeated region wherein the repeats are fixed. The number of such repeats varies from 1 to 40."
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
which is the codon for Pro or Ala wherein Pro = CCT, CCC, C
or CCG, and Ala = GCT, GCC, GCA, or GCG."
OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
which is the codon for Thr or Asn wherein Thr = ACT, ACC, AC
or ACG, and Asn = AAT or AAC."
OTHER INFORMATION: or ACG, and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
which is the codon for Pro or Ala wherein Pro = CCT, CCC, CC
or CCG, and Ala = GCT, GCC, GCA, or GCG."
OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147-21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537A-5

Query Match 42.6%; Score 1091; DB 2; Length 1867;
Best Local Similarity 76.5%; Pred. No. 86-75;
Matches 221; Conservative 5; Mismatches 61; Indels 2; Gaps 1;
QY 28 SPTPGKETSATORSSEVSTKKAASMTSSVLSHSPGSGSTTQGGDVTLPATEPA 67
Db 1570 STAPPAHGVTSAPDTRPPAGSTAPPAHGVTSAPDNKRPAGSTAPPAHGVTSAPDTRPP 1627
QY 88 SGSAAWGDVTSVETRPALGTTTPEPAHDVTSAPDNKRPAGSTAPPAHGVTSAPDTRPP 147
Db 1628 PGTAPPAHGVTSAPDTRPPAGSTAPPAHGVTSAPDNKRPAGSTAPPAHGVTSAPDTRPP 1687
QY 148 PGTAPPAHGVTSAPDTRPPAGSTAPPAHGVTSAPDNKRPAGSTAPPAHGVTSAPDTRPP 207
Db 1688 PGTAPPAHGVTSAPDTRPPAGSTAPPAHGVTSAPDNKRPAGSTAPPAHGVTSAPDTRPP 1747
QY 208 SASLTVANGTSARATTPPAKSTPPSIPSHSDPTTLASHTKTDASSTHSTVPLTS 267
Db 1748 SASLTVANGTSARATTPPAKSTPPSIPSHSDPTTLASHTKTDASSTHSTVPLTS 1807
QY 268 SNHSTPQSTGVSEFFLSPHISNLOPNSSLEPSTDYQELORDISEMPLQYKGGFL 316
Db 1808 SNHSTPQSTGVSEFFLSPHISNLOPNSSLEPSTDYQELORDISEMPLQYKGGFL 1856
RESULT 5
US-09-083-116-5
Sequence 5, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre

Thu May. 8 16:14:00 2003

us-09-658-621b-2.ra1

Page 5

APPLICANT: KIENY, Marie-Paule
APPLICANT: LARIE, Richard
APPLICANT: HARBERT, Robert
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22113-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE: 08-07-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
NAME/KEY: Peptide
LOCATION: 128-1727
OTHER INFORMATION: /note= "The amino acids spanning
128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:
NAME/KEY: Peptide
LOCATION: 128-1727
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-5
Query Match 42.6%; Score 1091; DB 4; Length 1867;
Best Local Similarity 76.5%; Pred. No. 8e-75;
Matches 221; Conservative 5; Mismatches 61; Indels 2; Gaps 1;
QY 28 SSTRPGGEKETSATORSSVSTETKQNAVMTSSVLSHSPSGSGSTTQODVTLAPYTEPA 87
DB 1570 STAPFAHGVTSAPDXRFXPGSTAPXAHGVTSAPDXRFXPGSTAPFAHGS--VTAPDXRFX 1627
QY 88 SGAATWQODVTSVPTRPALGSTTPPADVTSAPDKRPAFGSTAPPAHGVTSAPDTRPP 147
DB 1628 PGSTAPXAHGVTSAPDXRFXPGSTAPXAHGVTSAPDXRFXPGSTAPXAHGVTSAPDXRFX 1687
QY 148 PGSTAPPAHGVTSAPDTRPPAPGSTAPPAHGVTSAPDNRPALGSTAPPVHNVTSAGSASG 207
DB 1688 PGSTAPXAHGVTSAPDXRFXPGSTAPXAHGVTSAPDXRFXPGSTAPPAHNVTSAGSASG 1747
QY 208 SASLTVNNGTSANATTPASKSTPESIPSHGCTPTTLASHGKWDASTHSTVPLTS 267
DB 1748 SASLTVNNGTSANATTPASKSTPESIPSHGCTPTTLASHGKWDASTHSTVPLTS 1807
QY 268 SNGSTSPOLSTGVSPFFLSFHTSNLQFNLSLEDPTDYQELORISFM 316
DB 1808 SNGSTSPOLSTGVSPFFLSFHTSNLQFNLSLEDPTDYQELORISFM 1856
RESULT 6
US-09-134-916A-5
Sequence 5, Application US/09134916A
Patent No. 6328956
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LARIE, Richard
APPLICANT: HARBERT, Robert
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22113-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,916A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576

Thu May 8 16:14:00 2003

us-09-658-621b-2.rat

Page 6

```

; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 128..1727
; OTHER INFORMATION: /note= "The amino acids spanning
; OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
; OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
; OTHER INFORMATION: repeats varies from 1 to 40."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 134
; OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG,
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 144
; OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG,
; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 147
; OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG,
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
; NAME/KEY: Peptide
; LOCATION: 1..21
; OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
; OTHER INFORMATION: 21 amino acid precursor sequence."
; US-09-134-916A-5

Query Match 42.6%; Score 1091; DB 4; Length 1867;
Best Local Similarity 76.5%; Pred. No. 8e-75;
Matches 221; Conservative 5; Mismatches 61; Indels 2; Gaps 1;

QY 28 SSTPGGKERTATQSSVPSSTKNAVSMTSVLSHSPGSGSTTGQGVTLAPTEPA 87
DB 1570 STAPKAGVTSAPDXRPRXGTAAPKAGVTSAPDXRPRXGTAAPKAGVTSAPDXRPRX 1627
QY 88 SGGAATWQODVTVTPTRPALGSTTPRADVTSAPDXRPRXGTAAPKAGVTSAPDXRPRX 147
DB 1628 PGSTAPKAGVTSAPDXRPRXGTAAPKAGVTSAPDXRPRXGTAAPKAGVTSAPDXRPRX 1667
QY 148 PGSTAPKAGVTSAPDXRPRXGTAAPKAGVTSAPDXRPRXGTAAPKAGVTSAPDXRPRX 207
DB 1688 PGSTAPKAGVTSAPDXRPRXGTAAPKAGVTSAPDXRPRXGTAAPKAGVTSAPDXRPRX 1747
QY 208 SASTLVNAGTSARATTPPAKSTPPSPSHSDPTTLAASHKTIDASTHSTVPLPLTS 267
DB 1748 SASTLVNAGTSARATTPPAKSTPPSPSHSDPTTLAASHKTIDASTHSTVPLPLTS 1807
QY 268 SNHSTSPOLSTGVSPFFLSFHSINLQFNLSLDPSTDYVLOLORDISEM 316
DB 1808 SNHSTSPOLSTGVSPFFLSFHSINLQFNLSLDPSTDYVLOLORDISEM 1856

RESULT 7
```

```

US-08-783-774-2
; Sequence 2, Application US/08783774
; Patent No. 6054130
; GENERAL INFORMATION:
; APPLICANT: Spaece, Richard
; TITLE OF INVENTION: NON-SPLICING VARIANTS OF
; TITLE OF INVENTION: GP350/220
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; COMPUTER FILE: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PASTSIO Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/783,774
; FILING DATE: 15-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7682-037
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-669-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-783-774-2

Query Match 9.2%; Score 234.5; DB 3; Length 907;
Best Local Similarity 23.1%; Pred. No. 9.3e-10;
Matches 123; Conservative 62; Mismatches 194; Indels 153; Gaps 22;

QY 22 TGSCHAST-----PGGEKERTATQSSVPSSTKNAVSMTSVLSHSPGSGSTT 73
DB 444 TTTGCPSSHTVPTNLTPASTGPTVST--ADVTSPTPAGTTSGASVTPSPMDNGTES 501
QY 74 QGQDVT--LAPATEPASGAATWQODVTSVP-----VTRPALGSTPEPAH 116
DB 502 KAPMTSTSPVTTETPNATSPRAVTTTPPNATSPRAVTTTPPNATSPRLGKTSPSA 561
QY 117 DVTSAPD-NKPAFGSTAPRAHGVTSAPD--TRPPGSTAPRAHGVTSAPD--TRPAGSTAP 173
DB 562 VTTTPPMTSPFLGKTSPTSAVTTTPPNATSPFLGKTSPTSAVTTTPPNATSPFLGKTSPTSA 621
QY 174 PAHG-----VTSAPDNRPALGSTAPRAHGVTSASGSA-----SGSAST-- 212
DB 622 QANNTNHTAGTSPTPVVTSQPNATSAVTTGQ--HNITSSSTSSNKLAPSSNHTLSP 679
QY 213 -VHNGTSARATTPA-----SKGTPSPSPSHSDPTTLAASHKTIDASTHSTVPLPLTS 261
DB 680 TSDNSTHMPFLTSAHPTGEMNTQVTPASISTHVTSSSPERPQGTQAAGCGNNGSTS 739
QY 262 -----VPLPSSNHTSPOLSTGVSPFFLSFHSINLQFNLSL-----EDP 301
DB 740 TKPQSVVTVTKTPPNATSPFAQSKTAVPTVSTGKANSITGKHTTGARTSTER 799
QY 302 STDVYLOLORDISEMFLQIYKQAGFLGSLNKKRPQSVVVOVLTIAFREGTINVAHDVETOR 361
DB 800 TTDV-----GQSDTTPPRRYNATVYLPST----- 824
```

QY 362 NOYTEASRYNLITSDVSH--VPPPSAGAGVPGW-GIALVIV-CVVALAIY 417
 Db 825 ---SSKLRFRWFTSPPTTAQATVPVPTSQ-----FRPSNLSMLVLOMASLAVLTL 876
 QY 418 LIALAVGCCRRKRYGQLDIFPARDTHPMSEYFTYTHGRVPPSTDRSY 469
 Db 877 LVMADCAFR-----NLSISHT---YTPPYDAET 906

RESULT 8
 US-09-328-599A-1
 Sequence 1, Application US/09328599A
 Patent No. 6432679
 GENERAL INFORMATION:
 APPLICANT: MOND, James J. and Lees, Andrew
 TITLE OF INVENTION: Enhancement of B Cell Activation by
 TITLE OF INVENTION: Co-ligation of Receptors for Antigen and Complement C3d
 TITLE OF INVENTION: Using EBV gp350/220 or EBV gp350/220 Peptide Adjuvants
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
 ADDRESSEE: Dunner, L.L.P.
 STREET: 1300 I Street, N.W., Suite 700
 CITY: Washington
 STATE: DC
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/328,599A
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Forde, Jean B.
 REGISTRATION NUMBER: 32,984
 REFERENCE/DOCKET NUMBER: 04995, 6025-00000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)408-4000
 TELEFAX: (202)408-4000
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 907 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-328-599A-1

Query Match 9.2%; Score 234.5; DB 4; Length 907;
 Best Local Similarity 23.1%; Pred. No. 9.3e-10;
 Matches 123; Conservative 62; Mismatches 194; Indels 153; Gaps 22;
 QY 22 TSGHAST-----PGKEKTSATORSVPSSTKNAVMTSSVLSHSPSGSSST 73
 Db 444 TTGSPSTHVPTNLTPASTGPTVST--ADVTSPITAGTSGAPVTPSPMDNGTES 501
 QY 74 QGQDVT--LAPATEPASGSAATWGDVTSV-----VTRPALGSTTPPAH 116
 Db 502 KAPDMISSTSVTTTPNNISPTPAVITTPNNATSPPAATTPPNNATSPILAKISPTSA 561
 QY 117 DVTSPAD-NKPAGSTAPAGVTSAPD--TRPPGSTAPAPAGVTSAPD--TRAPGSTAP 173
 Db 562 VITPNNATSPILAKISPTSAVITTPNNATSPILAKISPTSAVITTPNNATSPILAKISPT 621
 QY 174 PANG-----VTSAPDNRPALGSTAPVNAHVTSASGSA-----SGSASTL-- 212
 Db 622 QANATNHTLGGTSPVVTSPQPKNATSAVTGO--NHTSSSTSSMSLRPSNPETLSFS 679

QY 213 -VANGTSAAATTPA-----SKTPPSISHSDDPTTLASHTKDSSTHST-- 261
 Db 680 TSNSTSHMPLTISAHPGTGNNITQVTPASISTHVTSSBPRGCTTSQASGPGSSPTS 739
 QY 262 -----VPLTSSNSTSPOLSTGVFFFLSFHISNLOFNSSL-----EDP 301
 Db 740 TRGEVNVTKGTTPQVATSPQAPSGQKAVPTVITSGKANSYTGKATTHGARTSTP 799
 QY 302 STDYQELQORDISEMFLQIKXQGFGLSNIKFRPGSVVOLTFLAFREGTINVDVETOF 361
 Db 800 TTDY-----GDSSTTPPRRYNATVTLPTST----- 824
 QY 362 NOYTEASRYNLITSDVSH--VPPPSAGAGVPGW-GIALVIV-CVVALAIY 417
 Db 825 ---SSKLRFRWFTSPPTTAQATVPVPTSQ-----FRPSNLSMLVLOMASLAVLTL 876
 QY 418 LIALAVGCCRRKRYGQLDIFPARDTHPMSEYFTYTHGRVPPSTDRSY 469
 Db 877 LVMADCAFR-----NLSISHT---YTPPYDAET 906

RESULT 9
 PCT-US95-04611A-19
 Sequence 19, Application PC/TUS9504611A
 GENERAL INFORMATION:
 APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
 TITLE OF INVENTION: Non Splicing Variants of gp350/220
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
 STREET: 5 Palo Alto Square
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/04611A
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/229,291
 FILING DATE: April 18, 1994
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Luann Caert
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: AVIR-003/00US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-843-5163
 TELEFAX: 415-857-0663
 TELEX: 380816 CooleyPA
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 907 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-04611A-19

Query Match 9.2%; Score 234.5; DB 5; Length 907;
 Best Local Similarity 23.1%; Pred. No. 9.3e-10;
 Matches 123; Conservative 62; Mismatches 194; Indels 153; Gaps 22;
 QY 22 TSGHAST-----PGKEKTSATORSVPSSTKNAVMTSSVLSHSPSGSSST 73
 Db 444 TTGSPSTHVPTNLTPASTGPTVST--ADVTSPITAGTSGAPVTPSPMDNGTES 501
 QY 74 QGQDVT--LAPATEPASGSAATWGDVTSV-----VTRPALGSTTPPAH 116

Thu May 8 16:14:00 2003

us-09-658-621b-2.ra1

Page 8

Db	502	KNDTSTSTSVTTTPPNNASPTPAVTTTPNNASTPAVTTTPPNNASTP	561
Oy	117	DVNSAPD--NKAPGSTPAPHGVTSAP--TPPSTSTAAHGVTSAPD--TPAGSGAP	173
Db	562	VTPPNNASTPPLTKSTPSTPAVTTTPNASTPPLTKSTPSTPAVTTTPPNNASTP	621
Oy	174	PAG--VTSAPDNRPALGSAAPVNNVTSAGSA--SGSAGL---	212
Db	622	QAAATHTLGGTSEPVPVTSQPKNASTAVTTCQ--NNITSSTSMGLRPSGAPPLTSPS	679
Oy	213	VHNGTSARATTPA-----SKTPPSIPIHSHDPTTLIAHSTKTPASGTHNSP-	261
Db	680	TSQNSTHMPPLTSTAHTGENTITVPAISLTHAVTSTSPBPAPGTTSAQSPGNSST	731
Oy	262	-----VPLTSTHNSSTPOUSTGVSEFFLSTHNSLQNSL-----	309
Db	740	TKGENTVTKTPPQNNITBPQAPSOQKNAVPVYTSIKANSSTGGHTTHGCASTSP	799
Oy	302	STPYVOELQORDISEMFLQYKGGFLGSLNKRPSGVVOLTALFABGFINHDOVQF	361
Db	800	TPDY-----GQDSTPPRRNATTPLPST-----	824
Oy	362	NOYKTEASRNLISIVGSVSH--VPPFSAQSGAGVPM--GIALTLV--CVLVAIAY	417
Db	825	-----SSKLPKRPTEFSPVTTAAATVAPVETSQ-----PAPSNLMLQWASIAVLITLL	876
Oy	418	LIALAVCCGRKNYQGLDIFPARDPYHMSSEYFTYTHHGRVYPPSSSDRAFY	469
Db	877	LIVAMDCAFR-----NLSTGSH-----VTPPYDDAETV	906

```

1      RESULT 10
2      US-09-556-706B-2
3      ; Sequence 2, Application US/09556706B
4      Patent No. 6458364
5      GENERAL INFORMATION:
6      APPLICANT: Spaeete, Richard
7      APPLICANT: Jackman, Winthrop
8      TITLE OF INVENTION: NON-SPLICING VARIANTS OF GP350/220
9      FILE REFERENCE: US/2000-04-24
10     CURRENT APPLICATION NUMBER: US/09/556, 706B
11     PRIOR APPLICATION NUMBER: 08/783, 774
12     PRIOR FILING DATE: 1997-01-15
13     PRIOR APPLICATION NUMBER: 08/229, 291
14     PRIOR FILING DATE: 1994-04-18
15     NUMBER OF SEQ ID NOS: 19
16     SOFTWARE: PatentIn version 3.0
17     SEQ ID NO: 2
18     LENGTH: 878
19     TYPE: PRT
20     ORGANISM: Virus
21     FEATURE:
22     OTHER INFORMATION: gp350
23     US-09-556-706B-2
24
25     Query Match      8 9%  Score 228; DB 4; Length 878;
26     Best Local Similarity 22.8%; PctIdent 60; Mismatches 187; Indels 152; Gaps
27     Matches 118; Conservative
28
29     Db
30
31     Qy 22 TGSCHASR-----PCGEKETASITQSSVPSRKNAAVSMTSYLSHSPSSSST 73
32     | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
33     Db 444 TTGDLPSSTHPTMLTAPASTGPTVSF--ADVTSPTAGTTSAGSPYTPSPSPMDNGTSS 501
34
35     Qy 74 QGDGV--LAPATEPAGSGAATMGQDVTSVP-----VTPEALGATTPPAH 116
36     | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
37     Db 502 KAPMTSSSTSPVTPFPNATSTPFAVTTTPFNATSPFAVTTTPFNATSPFLGTSPTSA 561
38
39     Qy 117 DVTSAPD--NKAPAGSTAPPAHGVTSAPD--TRPPGSTAPPAHGVTSAPD--TRPAGSTAP 173
40     | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
41     Db 562 VTTTPNATSPFLGKTSPTSAVTTTPFNATSPFLGKTSPTSAVTTTPNAGTSPVETSP 621

```

QY	174	PARGVTSADNRPALGSLTP-	PYANVTASQSGAGASGLVHNTQTSARLTPPSK--	229
Db	622	QANATVM-----	TLOGTSPRPV--VTSPQKQATSAVTTQGHNPSPSPETLSPSTDS	673
QY	230	-----TPSPISGHSADTPPLTASSTVTASSTHNSP-----	-----VPELTSS	268
Db	674	TSNMGSENIITQVPAISITSHVSTSSPEPRGQITSSQASQSPNSSTSTKGEVNTKQTP		733
QY	269	NHSTPQSLSTVSGEFFLSHLSINLQNSL-----	-----EDPSTDYQELQRISE	315
Db	734	QANHSQAPSQOKTAPVPTVTSIGKANSITGKQKHTTGHCATSTETPTDY-----		783
QY	316	MFQIQIKQSGSLGSLNKKRFGSVQQLTLAFRQGTINHDVEIQNSQVTKTAASRYNLT		375
Db	784	-----CGPSTTPPRRYNATLTLPSPH-----	-----SSKLAPRHPPT	815
QY	376	ISDVSVSH--VPPFPGAGGAGVGM-GIALVLVVCVLALATVILALAVCCERRKNG		432
Db	816	SPRYTTAQTATVPVPPTSQ--PFSNLSMLTV-----	-----LILLVWADCAFRF--	858
QY	433	QLDIPFARDTYHMSERYPTVTHGRVVPSPSTDSRY		469
Db	859	-----MLSTST-----	-----VTPPYDALETU	877

```

RESULT 11
US-09-165-239A-4
Sequence 4, Application US/09165239A
Patent No. 6344554
GENERAL INFORMATION:
APPLICANT: JOHNSON, ALEXANDER
APPLICANT: BRAUN, BERNHARD R
TITLE OF INVENTION: POLINUCLEOTIDE SEQUENCES FROM CANDIDA
FILE OF INVENTION: GERMAN ENCODING POLYPEPTIDES ASSOCIATED WITH FILAMENTOUS
TITLE OF INVENTION: CRYPT
FILE REFERENCE: 220022000700
CURRENT APPLICATION NUMBER: US/09/165,239A
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/068,065
PRIOR FILING DATE: 1997-12-18
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 750
TYPE: PRT
ORGANISM: Candida albicans
US-09-165-239A-4

Query Match
Best Local Similarity 28.1%; Score 207.5; DB 4; Length 750;
Matches 98; Conservative 46; Mismatches 116; Indels 133; Gaps 16;

Dh 240 TSKRTLTGLGLGLTKKTKGCGDTLQDFCMDCQCDTSPSTITVST--SSAIVSPSSAP 297
Qy 56 MTSVLSHSPGSGSGSTGQDYL---AAATPAGSAAATMGQIVTSVPTRPALST 111
Dh 228 STTVTVSSSVTPSSSVPTTIVTVSSVETTPSSAPPTTIVTVSSVTFPE--SS 355
Qy 112 TPRAHVTSADKKAPAGSTAPRAHSVTSAPDRPPSS---TPRAH--- 156
Dh 356 APTTPSSAPSS--VPSSAPPTTPSSAPSSVPSSAPETETETTPTHLTTTQI 414
Qy 157 -----GVTSAPT 164
Dh 415 TVIVTVSCSNNAKSTEVTVGVVVVTSSEDTIVTTECPLETTTPVSSVSDTSTVAPET 474
Qy 165 RP---APGSTAPRAHVTSAPD--NRALAGSTAPPVYNTVTS---GSAQGSAST-- 211
Dh 475 TPSTAPSSAP---SSAPSSAPVETVETPVSIVTQSGTIVTVTSSCSNNAKSEAK 530
Qy 212 -----LVHNGTSARATTPPASKTP-----PSIP-----SHSDPTPTL 245

```

Thu May 8 16:14:00 2003

us-09-658-621b-2.ra1

Page 9

Db 531 VTTGVVVTSEDTVTTCPLTETTPATESAPATESVATESAPVAPESSAPGTE 590

Qy 246 ASHSTKTDASTHSTVPLTSSN---HSTSP 274

Db 591 TAPATES-APATESSPVAPGTESSVAPESSAP 622

RESULT 12

US-08-099-354-1

; Sequence 1, Application US/08099354
; Patent No. 5746144
; GENERAL INFORMATION:
; APPLICANT: FINN, OLIVERA J.
; APPLICANT: FONTENOT, J. D.
; APPLICANT: MONTELLARO, RONALD C.
; TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
; TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DAREY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/099,354
; FILING DATE: 30-JUL-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STRILLA, GEORGE W.
; REGISTRATION NUMBER: 18221
; REFERENCE/DOCKET NUMBER: 6137/202246
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3536
; TELEFAX: 202-822-0944
; TELETYPE: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-099-354-1

Query Match 8.0%; Score 205; DB 1; Length 40;
Best Local Similarity 95.0%; Pred. No. 2.1e-09;
Matches 38; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 141 APDTRPPGSTAPAPAGVTSAPDTRPAPGSTAPAPAGVTS 180
Db 1 APDTRPAGSTAPAPAGVTSAPDTRPAPGSTAPAPAGVTS 40

RESULT 13

US-08-288-059-7
; Sequence 7, Application US/08288059
; Patent No. 5827666
; GENERAL INFORMATION:
; APPLICANT: FINN, OLIVERA J.
; APPLICANT: FONTENOT, J. D.
; APPLICANT: MONTELLARO, RONALD C.
; TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
; TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ATTORNEY/AGENT INFORMATION:

; ADDRESSEE: CUSHMAN, DAREY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/288,059
; FILING DATE: 08-AUG-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: CHAPIN, MARILAN K.
; REGISTRATION NUMBER: 35,843
; REFERENCE/DOCKET NUMBER: 61137/205204
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3711
; TELEFAX: 202-822-0944
; TELETYPE: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-288-059-7

Query Match 8.0%; Score 205; DB 2; Length 40;
Best Local Similarity 95.0%; Pred. No. 2.1e-09;
Matches 38; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 141 APDTRPPGSTAPAPAGVTSAPDTRPAPGSTAPAPAGVTS 180
Db 1 APDTRPAGSTAPAPAGVTSAPDTRPAPGSTAPAPAGVTS 40

RESULT 14
US-08-911-393-2
; Sequence 2, Application US/08911393
; Patent No. 6323008
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, Marc
; APPLICANT: BARKER, William A.
; APPLICANT: BARKER, David J.
; APPLICANT: ZOPE, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; TITLE OF INVENTION: STABILIZED SACCHARIDES IN A DAIRY SOURCE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,393
; FILING DATE: 14-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-032-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909030
TELEFAX: (212)8669741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1060 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-911-393-2

Query Match 8.0%; Score 205; DB 4; Length 1060;
Best Local Similarity 27.9%; Pred. No. 2,1e-07;
Matches 88; Conservative 31; Mismatches 136; Indels 60; Gaps 16;

QY 16 TVTITVYSGHASTPGGEKETATGRSSVSESTKNAVMTSSVLSHSPGSGSTTQG 75
DB 688 TPSTPADSSAR--STPSTPADSSAHSSTPAPGDNGAHSSTPSTPQDSSAHSSTPADNGA 745
QY 76 QDVTLAPATEPAPSGSAAT---WGQDVTSVPYTRPALGTTTP---AHDVTSAPDNKPA 127
DB 746 HSTPSAPADSNHSTSTPSTPADNGAHSSTPSTPADNGAHSSTPSTPADNGAHSSTPSTPADNGA 805
QY 128 PGSTAPP---AHGVTSAP-----DTRPPPGS-----TAPA---AHGVTSAP----- 162
DB 806 HSTPSTPADNGAHSSTPAPADSNHSTPSTPQDNGAHSSTPAPADSNHSTPSTPADSSA 865
QY 163 -DTRAPPGS-----TAP---PAHGVTSAPDNKRALGSTRPP---VAVVTSAGSGASG 207
DB 866 HSTPSAPGNGAHSSTPAPADSNHSTPAPGDNGAHSSTPAPADNGAHSSTPAPADNGAHSSTPAPADNGA 925
QY 208 SASLTVHNGTSAARAT--TTPASKS--TPPSIPSH--HS--DTPTLASHSTKT--DAAST 257
DB 926 HSTPSTPADSSAHSSTPSTPADSSAHSSTPAPGDNGAHSSTPAPADNGAHSSTPAPADSSAHSSTPSTPQDSSA 985
QY 258 HHSTVPEPLTSSNHST 272
DB 986 HSTPSAPADSSAHSST 1000

RESULT 15

US-09-134-001C-4463
Sequence 4463 Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4463
LENGTH: 2137
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 7.7%; Score 198; DB 4; Length 2137;
Best Local Similarity 23.9%; Pred. No. 1,9e-06;
Matches 108; Conservative 71; Mismatches 181; Indels 92; Gaps 15;

QY 19 TVTITVYSGHASTPGGEKETATGRSSVSESTKNAVMTSSVLSHSPG----- 67

DB 1699 TSLSDSTSTSESGSTSTSESDSDASSTSLSESTSTSLSDSTSTSTSDASSTSMVSDS 1758
QY 68 -----SGSSTTQGVDTLAPATEP-----SGSAATWQDVTSVPYTRPALGSTT 112
DB 1759 NRASTSLSDSTSVSDSTASSTSEASSTSESTSTASSTSLSESTSVSDSTSTST 1818
QY 113 PPAHDVTSAPDNKPAAGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPA-----P 168
DB 1819 DSASTSTSESDNSR--STSLSESTSVSDS--TSASTSASASTSVSDSNASSTSL 1874
QY 169 GSTAPPAHGVTSAPDNKPAAGSTAPPAHGVTSAGSAGSASTLVHNGTS-----ARATT 223
DB 1875 GSTSTSVSDSTST--STASASTSTSESDSDASTSLGSGSTSTSLSDSTSTSTSDASST 1932
QY 224 TPAKSTPPSPISPHSDPTTLAHSSTKTDASTHSTVPLTSSNHSTSPQLSTGVSAFF 283
DB 1933 TSEASTSTSESDNSR--STSLSESTSVSDSTSTSTSEASSTSTSEASSTSTSESTSVS-- 1990
QY 284 FLSPHISNLQFNSLFDPSDTPYQELORDISEMFLQ---LYKQGFILGLSNIKFRPGSV 340
DB 1991 -----EASSSTSLSDSS--TSTSMSTSEFTSQSPINSQSFIDGS-----L 2030
QY 341 VQTLAERAGTINV-----HDVEYQFNQYKTEASGRYLTISDVSVSH----- 383
DB 2031 SEDTIVYQSKNTMTMLNKTGKDYDLOEQRGYTDSE---QHNETOQNMQDNHNNLDLHQN 2087
QY 384 -----VPPFSAQSGAGVPGWGIALLVLCV 409
DB 2088 RLQDKRVKQPTGEGDGVVSNGFIVAVAVLAI 2119

Search completed: May 1, 2003, 23:14:26
CPU time : 34.5357 secs

D6	241	TPPTTASHSTYTDASSHTSHSTVPPLTSSNNSTSPOLDSTGVSFPPLSHSISNLOQFNSSLEED	300
QY	301	PSSTDYOBELORDISSEMFLOIYKQGGFGLANIFRRGSAVVQULTARBEOTINVADVETQ	360
D6	301	PSSTDYOBELORDISSEMFLOIYKQGGFGLANIFRRGSAVVQULTARBEOTINVADVETQ	360
QY	361	FMQYTBKASRYNLTISDVSVEHVPFPFSAQSAQAGVPGMGIALVLVLCVYVLAATVYLI	420
D6	361	FMQYTBKASRYNLTISDVSVEHVPFPFSAQSAQAGVPGMGIALVLVLCVYVLAATVYLI	420
QY	421	LAVCCCRKKNYGOGLDIFPPARDYVHMSREPYTHGRVYVPBSTDRSPYEKVSACNCGSS	480
D6	421	LAVCCCRKKNYGOGLDIFPPARDYVHMSREPYTHGRVYVPBSTDRSPYEKVSACNCGSS	480
QY	481	LSATYPAVAATSAATL	495
D6	481	LSATYPAVAATSAATL	495

```

RESULT 2
US-09-658-621B-67
: Sequence 67, Application US/09568621B
GENERAL INFORMATION:
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Heukamp, Lukas Carl
APPLICANT: Offringa, Rienk
APPLICANT: Melief, Cornelis Johanna Maria
APPLICANT: Acres, Bruce
APPLICANT: Thomas, Mireille
TITLE OF INVENTION: MHC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT FILING DATE: 2000-09-08
CURRENT APPLICATION NUMBER: US/09/658, 621B
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187, 215
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: GB 992242.5
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn version 3.1
SEQ ID NO 67
:
: TYPE: PRT
: LENGTH: 475
:
: ORGANISM: Homo sapiens
US-09-658-621B-67

```

Query Match	Similarity	Score	2413; DB 5;	Length	475;
Best Local	94.26%;	Pred. No. 5,7e-97;			
Matches	473; Conservative	0; Mismatches	2; Indels	20; Gaps	1
Oy	1	MPGQSPFFLLILLTLVTLVVSGSHASSTPGCEXETATSOSSVPSSTENKNAVMSATSV	60		
Db	1	MPGQSPFFLLILLTLVTLVVSGSHASSTPGCEXETATSOSSVPSSTENKNAVMSATSV	60		
Oy	61	LSHSHPGSGSSSTTGQDPTLAPEPSSGSAATVGCQDVSVPVTPALGSGTPPADYTS	120		
Db	61	LSHSHPGSGSSSTTGQDPTLAPEPSSGSAATVGCQDVSVPVTPALGSGTPPADYTS	120		
Oy	121	ADPNKPAAGSTAPPAHGTSAAPTRPPBSTAPPAHGVTSAPDTPAAGTAAAPAGYTS	180		
Db	121	ADPNKPAAGSTAPPAHGTSAAPTRPPBSTAPPAHGVTSAPDTPAAGTAAAPAGYTS	180		
Oy	181	ADPNKPAAGSTAPPAHVTNYSAGSGAGSASTLVANGTSAATVTPAAGSTPPELSHSD	240		
Db	161	ADPNKPAAGSTAPPAHVTNYSAGSGAGSASTLVANGTSAATVTPAAGSTPPELSHSD	220		
Oy	241	TPFTTLASHSTKTDASTTHSTHSTVPLTSSNHSSTPOLSTGVSPFELSPHTLSNLFNSLED	300		
Db	221	TPFTTLASHSTKTDASTTHSTHSTVPLTSSNHSSTPOLSTGVSPFELSPHTLSNLFNSLED	280		
Oy	301	PSTDYDGEGRDISENFLQIKGGAGFLGSLNTRKPPSSVVALTLAAREGTINWHDVETQ	360		

Db	281	PSTDYQRLORDISEMFLQIYKGGFLGINSIKFRPGSVVUOLILAFREGTINHVDETQ	340
Qy	361	FMQYTEAASRYNLITISDVSYSHAPPEFSASQDAGVPQMGIALLVYVCVLVAIAIVYLIA	420
Db	341	FMQYTEAASRYNLITISDVSYSDVPEFSASQDAGVPMQGIALLVYVCVLVAIAIVYLIA	400
Qy	421	LAVCCRRKRYNGOLDLFPARDTYHPMESEYPTATHGRVYPPSSYDTRSPYEKYSAGNGSS	480
Db	401	LAVCCRRKRYNGOLDLFPARDTYHPMESEYPTATHGRVYPPSSYDTRSPYEKYSAGNGSS	460
Qy	481	LSYTNBAVAATSNL	495
Db	461	LSYTNBAVAATSNL	475

```

RESULT 3
US-10-348-119-309
Sequence 309, Application US/10348119
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASES
TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
FILE REFERENCE: D0185 NP
CURRENT APPLICATION NUMBER: US/10/348,119
CURRENT FILING DATE: 2003-01-17
PRIOR APPLICATION NUMBER: US 60/350,061
NUMBER OF SEQ ID NOS: 795
SOFTWARE: PatentIn version 3.2
SEQ ID NO 309
LENGTH: 475
TYPE: prt
ORGANISM: Homo sapiens
US-10-348-119-309

```

Query Match	94.28;	Score	2413;	DB	6;	Length	475;
Best Local Similarity	95.64;	Pred.	No. 5.7e-97;				
Matches	473;	Conservative	0;	Mismatches	2;	Indels	20;
							Gaps
Qy	1	MPGQSPFLILLLTLVTWVGSGIASTPGCEKEKSAATQKRSVSPSTSKNAVMTSSV	60				
Db	1	MPGQSPFLILLLTLVTWVGSGIASTPGCEKEKSAATQKRSVSPSTSKNAVMTSSV	60				
Qy	61	LSHSPPSSSSTTQGDVTLAPTPBAGSAAITGQDVTSVVTRPALSTTPPAHDVTS	120				
Db	61	LSHSPPSSSSTTQGDVTLAPTPBAGSAAITGQDVTSVVTRPALSTTPPAHDVTS	120				
Qy	121	ADPKKAPGSTAPPAHGVTSAPDRPGSTAPPAHGVTSAPDRPAGSTAPPAHGVTS	180				
Db	121	ADPKKAPGSTAP-----AGVTSAPDRPAGSTAPPAHGVTS	160				
Qy	181	ADNRPALGSTAPPHNTSAGSSAGSASITLYNCGSARATTPAKSPSPISGSHSD	240				
Db	161	ADNRPALGSTAPPHNTSAGSSAGSASITLYNCGSARATTPAKSPSPISGSHSD	220				
Qy	241	PTTTLASHKTKDASTHSNTPPLTSSNHSSTQCTGVSEFLSHISNTQPMSSLED	300				
Db	221	PTTTLASHKTKDASTHSNTPPLTSSNHSSTQCTGVSEFLSHISNTQPMSSLED	280				
Qy	301	PESTDYQELORDISEMTQITKCGGTLGSLNKRPGSSVYVULIARREGTINHDVETQ	360				
Db	281	PESTDYQELORDISEMTQITKCGGTLGSLNKRPGSSVYVULIARREGTINHDVETQ	340				
Qy	361	PNQKTEPASRNLTISVSIVSHVPPSPSAQSGAGVPGMGIALVLVCVTLVALAIVYLIA	420				
Db	341	PNQKTEPASRNLTISVSIVSHVPPSPSAQSGAGVPGMGIALVLVCVTLVALAIVYLIA	400				
Qy	421	LAVCCRRKNVQGLDIPFADTYHPMSEVPTVTHGRVYVPSSTDSPSEYKVSAGNGSSS	480				
Db	401	LAVCCRRKNVQGLDIPFADTYHPMSEVPTVTHGRVYVPSSTDSPSEYKVSAGNGSSS	460				

Thu May. 8 16:14:03 2003

us-09-658-621b-2.rapn

Page 3

QY 481 LSYNPAVAATSNL 495
DB 461 LSYNPAVAATSNL 475

RESULT 4

US-60-422-176-59
Sequence 59, Application US/60422176
GENERAL INFORMATION:
APPLICANT: Pharmacia Corporation
TITLE OF INVENTION: Bournier, Maureen J.
FILE REFERENCE: 01040 PR
CURRENT APPLICATION NUMBER: US/60/422,176
NUMBER OF SEQ ID NOS: 3114
SOFTWARE: PatentIn version 3.1
LENGTH: 475
TYPE: PRT
ORGANISM: homo sapiens
US-60-422-176-59

Query Match 94.2%; Score 2413; DB 7; Length 475;
Best Local Similarity 95.6%; Pred. No. 5,7e-97;
Matches 473; Conservative 0; Mismatches 2; Indels 20; Gaps 1;

QY 1 MPTGOSPFLLLLTLTVVTSGHASTPGGKETSATGROSSVPSSTREKNAVSMTSV 60
DB 1 MPTGOSPFLLLLTLTVVTSGHASTPGGKETSATGROSSVPSSTREKNAVSMTSV 60
QY 61 LSSHPGSGSTTGGQDVTLPATERPAGSAATWGDVTSVPTRPALGSTTPPAHDVTS 120
DB 61 LSSHPGSGSTTGGQDVTLPATERPAGSAATWGDVTSVPTRPALGSTTPPAHDVTS 120
QY 121 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPAAHGVTSAPDTRPAGSTAPPAHGVTS 180
DB 121 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPAAHGVTSAPDTRPAGSTAPPAHGVTS 180
QY 181 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPAAHGVTSAPDTRPAGSTAPPAHGVTS 240
DB 181 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPAAHGVTSAPDTRPAGSTAPPAHGVTS 240
QY 241 TPTTLASHTKTDASTHSTHSTVPLTSSNHSTSPOLSTGVSPFPLSPHISMLQNSSLD 300
DB 241 TPTTLASHTKTDASTHSTHSTVPLTSSNHSTSPOLSTGVSPFPLSPHISMLQNSSLD 300
QY 301 PSTDYOELORDISEMFLQIKOGGFLGSLNKKRPPGSAVVOULTAFREGTINHDVETQ 360
DB 301 PSTDYOELORDISEMFLQIKOGGFLGSLNKKRPPGSAVVOULTAFREGTINHDVETQ 360
QY 361 FNOYKTEAARVNLITSDVSVAHPPFPAAGAGVPMGIALVLCVLAIAIYLLA 420
DB 361 FNOYKTEAARVNLITSDVSVAHPPFPAAGAGVPMGIALVLCVLAIAIYLLA 420
QY 421 LAVGCRKRYGQLDIPPARDTYHPMSEYPTHTHGRVPPSSDPSYEKVSAGNGSS 480
DB 421 LAVGCRKRYGQLDIPPARDTYHPMSEYPTHTHGRVPPSSDPSYEKVSAGNGSS 480
QY 481 LSYNPAVAATSNL 495
DB 461 LSYNPAVAATSNL 475

RESULT 5
US-60-452-680-12835
Sequence 12835, Application US/60452680
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CL001450
CURRENT APPLICATION NUMBER: US/60/452,680
CURRENT FILING DATE: 2003-03-07
NUMBER OF SEQ ID NOS: 116213
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12835
LENGTH: 475
TYPE: PRT
ORGANISM: Homo sapiens
US-60-452-680-12835

Query Match 94.2%; Score 2413; DB 7; Length 475;
Best Local Similarity 95.6%; Pred. No. 5,7e-97;
Matches 473; Conservative 0; Mismatches 2; Indels 20; Gaps 1;

QY 1 MPTGOSPFLLLLTLTVVTSGHASTPGGKETSATGROSSVPSSTREKNAVSMTSV 60
DB 1 MPTGOSPFLLLLTLTVVTSGHASTPGGKETSATGROSSVPSSTREKNAVSMTSV 60
QY 61 LSSHPGSGSTTGGQDVTLPATERPAGSAATWGDVTSVPTRPALGSTTPPAHDVTS 120
DB 61 LSSHPGSGSTTGGQDVTLPATERPAGSAATWGDVTSVPTRPALGSTTPPAHDVTS 120
QY 121 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPAAHGVTSAPDTRPAGSTAPPAHGVTS 180
DB 121 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPAAHGVTSAPDTRPAGSTAPPAHGVTS 180
QY 181 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPAAHGVTSAPDTRPAGSTAPPAHGVTS 240
DB 181 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPAAHGVTSAPDTRPAGSTAPPAHGVTS 240
QY 241 TPTTLASHTKTDASTHSTHSTVPLTSSNHSTSPOLSTGVSPFPLSPHISMLQNSSLD 300
DB 241 TPTTLASHTKTDASTHSTHSTVPLTSSNHSTSPOLSTGVSPFPLSPHISMLQNSSLD 300
QY 301 PSTDYOELORDISEMFLQIKOGGFLGSLNKKRPPGSAVVOULTAFREGTINHDVETQ 360
DB 301 PSTDYOELORDISEMFLQIKOGGFLGSLNKKRPPGSAVVOULTAFREGTINHDVETQ 360
QY 361 FNOYKTEAARVNLITSDVSVAHPPFPAAGAGVPMGIALVLCVLAIAIYLLA 420
DB 361 FNOYKTEAARVNLITSDVSVAHPPFPAAGAGVPMGIALVLCVLAIAIYLLA 420
QY 421 LAVGCRKRYGQLDIPPARDTYHPMSEYPTHTHGRVPPSSDPSYEKVSAGNGSS 480
DB 421 LAVGCRKRYGQLDIPPARDTYHPMSEYPTHTHGRVPPSSDPSYEKVSAGNGSS 480
QY 481 LSYNPAVAATSNL 495
DB 461 LSYNPAVAATSNL 475

RESULT 6
US-60-453-135-8036
Sequence 8036, Application US/60453135
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001456
CURRENT APPLICATION NUMBER: US/60/453,135
NUMBER OF SEQ ID NOS: 82762
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8036
LENGTH: 475
TYPE: PRT
ORGANISM: Homo sapiens
US-60-453-135-8036

Query Match 94.2%; Score 2413; DB 7; Length 475;
Best Local Similarity 95.6%; Pred. No. 5,7e-97;

Thu May 8 16:14:03 2003

us-09-658-621b-2.rapn

Page 4

	Matches	4/3; Conservative	0; Mismatches	2; Indels	20; Gaps	1
Qy	1	MPGQGPFLILLTLVTVYTGSGHASTPGGKERTSANQSSVPSSTKMNVMSTSV	60			
Dh	1	MPGQGPFLILLTLVTVYTGSGHASTPGGKERTSANQSSVPSSTKMNVMSTSV	60			
Qy	61	LSHSHPGSGSSTQGDVTLAPATBPASGAATMGQDVSVPVTPALGSGTTPADYTS	120			
Dh	61	LSHSHPGSGSSTQGDVTLAPATBPASGAATMGQDVSVPVTPALGSGTTPADYTS	120			
Qy	121	ABDNKPAAGSTAPAGVTSAPDTRPPPGSTAPAAHVZSAADTRPALGSGTTPAGVTS	180			
Dh	121	ADNKKPAPGSTAP-----AHGVZSAADTRPALGSGTTPAGVTS	160			
Qy	181	ABDNKPAAGSTAPVHNVTASGSGASGASSTLVHNGTSAATTTAPSKSTPSPISPHSD	240			
Dh	161	ABDNKPAAGSTAPVHNVTASGSGASGASSTLVHNGTSAATTTAPSKSTPSPISPHSD	220			
Qy	241	TPFTLLASHSTKTDASTHSTVPEPLTSSNHSTSPQSLSTGVSPFLSPHISNLDENSTSD	300			
Dh	221	TPFTLLASHSTKTDASTHSTVPEPLTSSNHSTSPQSLSTGVSPFLSPHISNLDENSTSD	280			
Qy	301	PSTTYQELORISSEMTQIKYCGSLGHSNKRPESSVYVLTAPREGTINVHDVSTQ	360			
Dh	281	PSTTDYQELORISSEMTQIKYCGSLGHSNKRPESSVYVLTAPREGTINVHDVSTQ	340			
Qy	361	PNQYKTEAPBARNYLTISDVSSISDVPPFPFPGSGAGVGMQALVAVCYLALVALYTLA	420			
Dh	341	PNQYKTEAPBARNYLTISDVSSISDVPPFPFPGSGAGVGMQALVAVCYLALVALYTLA	400			
Qy	421	LAVCGCKRKNYQGLDIPARBTYHPMSKPYTHHNGKYPSPSTGRSAPYKRYGANGGSS	480			
Dh	401	LAVCGCKRKNYQGLDIPARBTYHPMSKPYTHHNGKYPSPSTGRSAPYKRYGANGGSS	460			
Qy	481	LSYTNPAVAATSANL	495			
Dh	461	LSYTNPAVAATSANL	475			

```

RESULT 7
US-60-453-050-8036
Sequence 8036, Application US/60453050
GENERAL INFORMATION:
APPLICANT: CARILLI, Michele
APPLICANT: LUKE, MAY
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CL001457
CURRENT FILING DATE: 2003-03-10
CURRENT FILING DATE: 2003-03-10
NUMBER OF SEQ. ID NOS: 82762
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ. ID NO 8036
LENGTH: 475
TYPE: PAT
US-60-453-050-8036
ORGANISM: Homo sapiens

Query Match      94.2%; Score 2413; DB 7; Length 475;
Best Local Similarity 95.6%; Pred. No. 5.7e-97;
Matches 473; Conservative 0; Mismatches 2; Indels 20; Gaps 1.

Qy      1  MTPGQSPFLLLLTLVTVVVGSGHASTPCGSKETASORSSVPESTRNAVSTSSV 60
Db      1  MTPGQSPFLLLLTLVTVVVGSGHASTPCGSKETASORSSVPESTRNAVSTSSV 60

Qy      61  LSHSPGSGSTGQODVTLPAPEBPASGAATGQDVTSVTPPALGSGTTPPAIDVTS 120
Db      61  LSHSPGSGSTGQODVTLPAPEBPASGAATGQDVTSVTPPALGSGTTPPAIDVTS 120

Qy      121  APDNKPAAGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTPAPGSTAPPAHGVTS 180
Db      121  APDNKPAAGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTPAPGSTAPPAHGVTS 180
Qy      121  APDNKPAAGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTPAPGSTAPPAHGVTS 160
Db      121  APDNKPAAGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTPAPGSTAPPAHGVTS 160

```

Qy	181	AANBPAIGSTAPPPYNNYTSAGSAGSAGSASTLVNHCNTABATYTPPASKSTPSPISPMHSD	2.40
Db	161	APDNKPALGSTAPFPYNNYTSAGSAGSAGSASTLVNHCNTABATYTPPASKSTPSPISPMHSD	2.20
Qy	241	TPPTTLASSTYTSASTHHSVAPPPLTSGNNHSTSPOLSTGVSFFPLASHLSNLOFMSLSD	3.00
Db	221	TPPTTLASSTYTSASTHHSVAPPPLTSGNNHSTSPOLSTGVSFFPLASHLSNLOFMSLSD	2.80
Qy	301	PSYDYOHOLODISSEMFLQIYKOGSPLGLSNIKEPPSSVVVOLLTPAREGTINNVHDETQ	3.60
Db	281	PSYDYOHOLODISSEMFLQIYKOGSPLGLSNIKEPPSSVVVOLLTPAREGTINNVHDETQ	3.40
Qy	361	PNQYTSABSNYLITSDYSNVDPPPPSAGSAGVCEMGJALNYLCYVALVYIYA	4.20
Db	341	PNQYTSABSNYLITSDYSNVDPPPPSAGSAGVCEMGJALNYLCYVALVYIYA	4.00
Qy	421	LAVOCRRKQYQGLDFPPARDYHPMSEYLYTHYGRVVPSSSTDSPEEKSASNGGSS	4.80
Db	401	LAVOCRRKQYQGLDFPPARDYHPMSEYLYTHYGRVVPSSSTDSPEEKSASNGGSS	4.60
Qy	481	LSYTNPAVAATSNL 495	
Db	461	LSYTNPAVAATSNL 475	

[illegible]

```
Db 161 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 240
Qy 165 ----- 164
Db 241 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 300
Qy 165 ----- 164
Db 301 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 360
Qy 165 ----- 164
Db 361 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 420
Qy 165 ----- 164
Db 421 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 480
Qy 165 ----- 164
Db 481 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 540
Qy 165 ----- 164
Db 541 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 600
Qy 165 ----- 164
Db 601 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 660
Qy 165 ----- 164
Db 661 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 720
Qy 165 ----- 164
Db 721 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 780
Qy 165 ----- 164
Db 781 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 840
Qy 165 ----- 164
Db 841 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 900
Qy 165 ----- 164
Db 901 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 960
Qy 165 ----- 164
Db 961 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 1020
Qy 165 ----- 164
Db 1021 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 1080
Qy 165 ----- 164
Db 1081 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 1140
Qy 165 ----- 164
Db 1141 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 1200
Qy 165 ----- 164
Db 1201 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 1260
Qy 165 ----- 164
```

```
PCT-IL02-00255-2
SEQUENCE 2 Application PC/IL0200255
GENERAL INFORMATION:
APPLICANT: Ramot University Authority for Applied Research
INVENTOR: KERSCHNER, Daniel
APPLICANT: YOBEL-LERNER, Merav
APPLICANT: SMORODINSKY, Nechama
TITLE OF INVENTION: Peptides and Antibodies to MUC 1 Proteins
FILE REFERENCE: P-3891-PC
CURRENT APPLICATION NUMBER: PCT/IL02/00255
PRIORITY FILING DATE: 2002-03-26
PRIORITY FILING DATE: 2001-03-29
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 1255
TYPE: PRT
ORGANISM: Homo sapiens
PCT-IL02-00255-2

Query Match 83.1%; Score 2130; DB 1; Length 1255;
Best Local Similarity 39.0%; Pred. No. 2.1e-84;
Matches 489; Conservative 1; Mismatches 5; Indels 760; Gaps 1;

Qy 1 MPTGQSPFFLLTLTLVTVYVSGGASSTPGGKETSATQSSVPSSTERKAVSMTSSV 60
Db 1 MPTGQSPFFLLTLTLVTVYVSGGASSTPGGKETSATQSSVPSSTERKAVSMTSSV 60
Qy 61 LSSHPGSGSSTTGQDVTTLAPATEPAGSAAITWQDVTSPVTRPALGSTTPPAHDVTS 120
Db 61 LSSHPGSGSSTTGQDVTTLAPATEPAGSAAITWQDVTSPVTRPALGSTTPPAHDVTS 120
Qy 121 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 164
Db 121 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 164
Qy 165 ----- 164
Db 165 ----- 164
Qy 181 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 240
Db 181 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 240
Qy 165 ----- 164
Db 241 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 300
Qy 165 ----- 164
Db 301 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 360
Qy 165 ----- 164
Db 361 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 420
Qy 165 ----- 164
Db 421 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 480
Qy 165 ----- 164
Db 481 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 540
Qy 165 ----- 164
Db 541 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 600
Qy 165 ----- 164
Db 601 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 660
Qy 165 ----- 164
Db 661 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 720
```

```

OY 165 ----- 164
DB 721 APDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 780
OY 165 ----- 164
DB 781 APDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 840
OY 165 ----- 164
DB 841 APDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 900
OY 165 ----- 200
DB 901 APDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 960
OY 201 ASGSASGSASTLVHNGTSARATTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 260
DB 961 ASGSASGSASTLVHNGTSARATTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 1020
OY 261 TYPPLTSSNMHSTSPQSTGVSPFFLSFHSINLOFNSSLEDPSTDYQELORDISEMFLQI 320
DB 1021 SVPLTSSNMHSTSPQSTGVSPFFLSFHSINLOFNSSLEDPSTDYQELORDISEMFLQI 1080
OY 321 YKGGFLGSLNKKRPPGSVVVOLTAFREBGTIVHGVETOPNOYKTEASRYNLITSDVS 380
DB 1081 YKGGFLGSLNKKRPPGSVVVOLTAFREBGTIVHGVETOPNOYKTEASRYNLITSDVS 1140
OY 381 VSHVPPPSAOSGAGVPGMGIALVLVCVVALATVYLAVALAOCORRONGOLDIPPAR 440
DB 1141 VSDVPPPSAOSGAGVPGMGIALVLVCVVALATVYLAVALAOCORRONGOLDIPPAR 1200
OY 441 DTVHPMSSEPTHTHGRVPPSTDRSPYKVSAGSGSSISYTNPAVAATSANL 495
DB 1201 DTVHPMSSEPTHTHGRVPPSTDRSPYKVSAGSGSSISYTNPAVAATSANL 1255

```

```

RESULT 10
US-60-452-680-12838
; Sequence 12838, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGIL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 12838
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-12838

```

```

Query Match 77.8%; Score 1992.5; DB 7; Length 400;
Best Local Similarity 80.6%; Pred. No. 6.6e-79; Indels 95; Gaps 2;
Matches 399; Conservative 0; Mismatches 1; Indels 95; Gaps 2;

```

```

OY 1 MTGCTGSPFFLLILLVTLVTVGSGHASTPGEKETSATQSSVSSSTERNKAVSMSSV 60
DB 1 MTGCTGSPFFLLILLVTLVTVGSGHASTPGEKETSATQSSVSSSTERNKAVSMSSV 60
OY 61 LSHSFGSGSSSTGGDVTLPATEPAGSAAATMGODVTSVPVTRPALGSTTPPAHGVTS 120
DB 61 LSHSFGSGSSSTGGDVTLPATEPAGSAAATMGODVTSVPVTRPALGSTTPPAHGVTS 120
OY 121 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 180
DB 121 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 180
OY 121 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 160
DB 121 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 160
OY 161 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 240
DB 161 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 240

```

```

DB 161 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 220
OY 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNMHSTSPQSTGVSPFFLSFHSINLOFNSSLEDP 300
DB 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNMHSTSPQSTGVSPFFLSFHSINLOFNSSLEDP 280
OY 301 PSTDYQELORDISEMFLQIYKGGFLGSLNKKRPPGSVVVOLTAFREBGTIVHGVETOP 360
DB 281 PSTDYQELORDISEMFLQIYKGGFLGSLNKKRPPGSVVVOLTAFREBGTIVHGVETOP 296
OY 361 FNOYKTEASRYNLITSDVSVPFPPSAOSGAGVPGMGIALVLVCVVALATVYLA 420
DB 297 ----- 325
OY 421 LANCORRONGOLDIPPARDTPHMSSEPTHTHGRVPPSTDRSPYKVSAGSGSSISYTNPAVAATSANL 480
DB 326 LANCORRONGOLDIPPARDTPHMSSEPTHTHGRVPPSTDRSPYKVSAGSGSSISYTNPAVAATSANL 385
OY 481 LSYTNPAVAATSANL 495
DB 386 LSYTNPAVAATSANL 400

```

```

RESULT 11
US-60-453-135-8039
; Sequence 8039, Application US/60453135
; GENERAL INFORMATION:
; APPLICANT: CARGIL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001456
; CURRENT APPLICATION NUMBER: US/60/453,135
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 8039
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-135-8039

```

```

Query Match 77.8%; Score 1992.5; DB 7; Length 400;
Best Local Similarity 80.6%; Pred. No. 6.6e-79; Indels 95; Gaps 2;
Matches 399; Conservative 0; Mismatches 1; Indels 95; Gaps 2;

```

```

OY 1 MTGCTGSPFFLLILLVTLVTVGSGHASTPGEKETSATQSSVSSSTERNKAVSMSSV 60
DB 1 MTGCTGSPFFLLILLVTLVTVGSGHASTPGEKETSATQSSVSSSTERNKAVSMSSV 60
OY 61 LSHSFGSGSSSTGGDVTLPATEPAGSAAATMGODVTSVPVTRPALGSTTPPAHGVTS 120
DB 61 LSHSFGSGSSSTGGDVTLPATEPAGSAAATMGODVTSVPVTRPALGSTTPPAHGVTS 120
OY 121 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 180
DB 121 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 180
OY 121 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 160
DB 121 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 160
OY 161 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 240
DB 161 APNKRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTSAPDTPRPAAGSTAPPAHGVTS 220
OY 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNMHSTSPQSTGVSPFFLSFHSINLOFNSSLEDP 300
DB 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNMHSTSPQSTGVSPFFLSFHSINLOFNSSLEDP 280
OY 301 PSTDYQELORDISEMFLQIYKGGFLGSLNKKRPPGSVVVOLTAFREBGTIVHGVETOP 360
DB 281 PSTDYQELORDISEMFLQIYKGGFLGSLNKKRPPGSVVVOLTAFREBGTIVHGVETOP 296
OY 361 FNOYKTEASRYNLITSDVSVPFPPSAOSGAGVPGMGIALVLVCVVALATVYLA 420
DB 361 FNOYKTEASRYNLITSDVSVPFPPSAOSGAGVPGMGIALVLVCVVALATVYLA 420

```

Thu May 8 16:14:03 2003

us-09-658-621b-2.rapu

Page 7

```
Db 297 -----SGAGVPGMGIALVLVVCVLAALIVYLLA 325
Qy 421 LAVOCRRKRYGQGLDIFPARDTYHPMSSEPTHTHGRYVPPSSTRSPYEVKVSAGNGSS 480
Db 326 LAVOCRRKRYGQGLDIFPARDTYHPMSSEPTHTHGRYVPPSSTRSPYEVKVSAGNGSS 385
Qy 481 LSTYNPAVAATSNL 495
Db 386 LSTYNPAVAATSNL 400

RESULT 12
US-60-453-050-8039
; Sequence 8039, Application US/60453050
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01457
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8039
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-050-8039

Query Match 77.8%; Score 1992.5; DB 7; Length 400;
Beat Local Similarity 80.6%; Pred. No. 6.6e-79;
Matches 399; Conservative 0; Mismatches 1; Indels 95; Gaps 2;

Qy 1 MTPGTGSPFFLLLLTLVLTAVTSGHASTPGGKETSATORSVSPSTEKNAVMTSSV 60
Db 1 MTPGTGSPFFLLLLTLVLTAVTSGHASTPGGKETSATORSVSPSTEKNAVMTSSV 60
Qy 61 LSHSPGSGSSTTGGQDVTLPATPEPASGSAATWGQDVTSPVTRPALGSTPPADVTS 120
Db 61 LSHSPGSGSSTTGGQDVTLPATPEPASGSAATWGQDVTSPVTRPALGSTPPADVTS 120
Qy 61 LSHSPGSGSSTTGGQDVTLPATPEPASGSAATWGQDVTSPVTRPALGSTPPADVTS 120
Db 61 LSHSPGSGSSTTGGQDVTLPATPEPASGSAATWGQDVTSPVTRPALGSTPPADVTS 120
Qy 121 APDKKPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
Db 121 APDKKPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
Qy 121 APDKKPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
Db 121 APDKKPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
Qy 181 APDKRPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
Db 181 APDKRPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
Qy 161 APDKRPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 220
Db 161 APDKRPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 220
Qy 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHTNLQPNLSLED 300
Db 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHTNLQPNLSLED 300
Qy 221 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHTNLQPNLSLED 280
Db 221 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHTNLQPNLSLED 280
Qy 301 PSTDYGOELORDISEMFLQIYKGGFLGSLNFKRPGSVVQLTLPAREGTINWHDVETQ 360
Db 301 PSTDYGOELORDISEMFLQIYKGGFLGSLNFKRPGSVVQLTLPAREGTINWHDVETQ 360
Qy 281 PSTDYGOELORDISEMFLQIYKGGFLGSLNFKRPGSVVQLTLPAREGTINWHDVETQ 296
Db 281 PSTDYGOELORDISEMFLQIYKGGFLGSLNFKRPGSVVQLTLPAREGTINWHDVETQ 296
Qy 361 FNOYKTEASRYNLITDVSVSHVPPPSAQSAGVPGMGIALVLVVCVLAALIVYLLA 420
Db 361 FNOYKTEASRYNLITDVSVSHVPPPSAQSAGVPGMGIALVLVVCVLAALIVYLLA 420
Qy 297 -----SGAGVPGMGIALVLVVCVLAALIVYLLA 325
Db 297 -----SGAGVPGMGIALVLVVCVLAALIVYLLA 325
Qy 421 LAVOCRRKRYGQGLDIFPARDTYHPMSSEPTHTHGRYVPPSSTRSPYEVKVSAGNGSS 480
Db 421 LAVOCRRKRYGQGLDIFPARDTYHPMSSEPTHTHGRYVPPSSTRSPYEVKVSAGNGSS 480
Qy 326 LAVOCRRKRYGQGLDIFPARDTYHPMSSEPTHTHGRYVPPSSTRSPYEVKVSAGNGSS 385
Db 326 LAVOCRRKRYGQGLDIFPARDTYHPMSSEPTHTHGRYVPPSSTRSPYEVKVSAGNGSS 385
Qy 481 LSTYNPAVAATSNL 495
Db 481 LSTYNPAVAATSNL 400
Qy 386 LSTYNPAVAATSNL 400
Db 386 LSTYNPAVAATSNL 400

RESULT 13
US-60-452-680-12837
```

```
; Sequence 12837, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01450
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12837
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-12837

Query Match 71.5%; Score 1831.5; DB 7; Length 370;
Beat Local Similarity 74.5%; Pred. No. 5.4e-72;
Matches 369; Conservative 0; Mismatches 1; Indels 125; Gaps 2;

Qy 1 MTPGTGSPFFLLLLTLVLTAVTSGHASTPGGKETSATORSVSPSTEKNAVMTSSV 60
Db 1 MTPGTGSPFFLLLLTLVLTAVTSGHASTPGGKETSATORSVSPSTEKNAVMTSSV 60
Qy 61 LSHSPGSGSSTTGGQDVTLPATPEPASGSAATWGQDVTSPVTRPALGSTPPADVTS 120
Db 61 LSHSPGSGSSTTGGQDVTLPATPEPASGSAATWGQDVTSPVTRPALGSTPPADVTS 120
Qy 61 LSHSPGSGSSTTGGQDVTLPATPEPASGSAATWGQDVTSPVTRPALGSTPPADVTS 120
Db 61 LSHSPGSGSSTTGGQDVTLPATPEPASGSAATWGQDVTSPVTRPALGSTPPADVTS 120
Qy 121 APDKKPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
Db 121 APDKKPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
Qy 121 APDKKPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
Db 121 APDKKPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
Qy 181 APDKRPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
Db 181 APDKRPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
Qy 161 APDKRPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 220
Db 161 APDKRPAPOSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 220
Qy 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHTNLQPNLSLED 300
Db 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHTNLQPNLSLED 300
Qy 221 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHTNLQPNLSLED 280
Db 221 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHTNLQPNLSLED 280
Qy 301 PSTDYGOELORDISEMFLQIYKGGFLGSLNFKRPGSVVQLTLPAREGTINWHDVETQ 360
Db 301 PSTDYGOELORDISEMFLQIYKGGFLGSLNFKRPGSVVQLTLPAREGTINWHDVETQ 360
Qy 281 PSTDYGOELORDISEMFLQIYKGGFLGSLNFKRPGSVVQLTLPAREGTINWHDVETQ 296
Db 281 PSTDYGOELORDISEMFLQIYKGGFLGSLNFKRPGSVVQLTLPAREGTINWHDVETQ 296
Qy 361 FNOYKTEASRYNLITDVSVSHVPPPSAQSAGVPGMGIALVLVVCVLAALIVYLLA 420
Db 361 FNOYKTEASRYNLITDVSVSHVPPPSAQSAGVPGMGIALVLVVCVLAALIVYLLA 420
Qy 297 -----296
Db 297 -----296
Qy 421 LAVOCRRKRYGQGLDIFPARDTYHPMSSEPTHTHGRYVPPSSTRSPYEVKVSAGNGSS 480
Db 421 LAVOCRRKRYGQGLDIFPARDTYHPMSSEPTHTHGRYVPPSSTRSPYEVKVSAGNGSS 480
Qy 297 -AVOCRRKRYGQGLDIFPARDTYHPMSSEPTHTHGRYVPPSSTRSPYEVKVSAGNGSS 355
Db 297 -AVOCRRKRYGQGLDIFPARDTYHPMSSEPTHTHGRYVPPSSTRSPYEVKVSAGNGSS 355
Qy 481 LSTYNPAVAATSNL 495
Db 481 LSTYNPAVAATSNL 370
Qy 356 LSTYNPAVAATSNL 370
Db 356 LSTYNPAVAATSNL 370

RESULT 14
US-60-453-135-8038
; Sequence 8038, Application US/60453135
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: IAKUBOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01456
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8038
; LENGTH: 370
```

Thu May 8 16:14:03 2003

us-09-658-621b-2.rapn

Page 8

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-135-8038

Query Match
Best Local Similarity 71.5%; Score 1831.5; DB 7; Length 370;
Matches 369; Conservative 0; Mismatches 1; Indels 125; Gaps

OY 1 MPPTGSPFFLLILLTLVTVVTSAGSASTPGCKETSATROSSVPSSTKNAVMSSTSV 60
Db 1 MPPTGSPFFLLILLTLVTVVTSAGSASTPGCKETSATROSSVPSSTKNAVMSSTSV 60
OY 61 LSSHPSGSGSSTTQODVTLAPATERPASGAATWGOVTVSVPTPALGSTTPPAHDVTS 120
Db 61 LSSHPSGSGSSTTQODVTLAPATERPASGAATWGOVTVSVPTPALGSTTPPAHDVTS 120
OY 121 APDNKPAVGTAPPAHVGTAPADTPRPPGSTAPPAHVGTAPADTPRPPGSTAPPAHVGT 180
Db 121 APDNKPAVGTAPPAHVGTAPADTPRPPGSTAPPAHVGTAPADTPRPPGSTAPPAHVGT 180
OY 181 APDNKPAVGTAPPAHVGTAPADTPRPPGSTAPPAHVGTAPADTPRPPGSTAPPAHVGT 240
Db 181 APDNKPAVGTAPPAHVGTAPADTPRPPGSTAPPAHVGTAPADTPRPPGSTAPPAHVGT 240
OY 241 PPTTAASTKTDASTTHNSVTPPLTSSNHSSTQSTGVSFFLSPIINLOPMSILED 300
Db 241 PPTTAASTKTDASTTHNSVTPPLTSSNHSSTQSTGVSFFLSPIINLOPMSILED 300
OY 221 PPTTAASTKTDASTTHNSVTPPLTSSNHSSTQSTGVSFFLSPIINLOPMSILED 280
Db 221 PPTTAASTKTDASTTHNSVTPPLTSSNHSSTQSTGVSFFLSPIINLOPMSILED 280
OY 301 PSTDYDELORDISEMFLQIKKGGSLGSLNKKRPSSVVQGLAPREGTIVHDVETQ 360
Db 281 PSTDYDELORDISEM----- 296
OY 361 PNOYKTBASRYVLTISDVSVHPFPFSAAGAGVPCMGIALVLVYVALATVYLLA 420
Db 297 ----- 296
OY 421 LAVQCGRRKRVYQGLDIPARPTVHNSKYPCTHGHGVNPPSTPRAPRYKVASGSGSS 480
Db 297 -AVCGRRKRVYQGLDIPARPTVHNSKYPCTHGHGVNPPSTPRAPRYKVASGSGSS 355
OY 481 LSTYTPAAVTAASNL 495
Db 356 LSTYTPAAVTAASNL 370

RESULT 15
US-60-453-050-8038
; Sequence 8038 Application US/60453050
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: JUNE
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: STENOISIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: 010145
; CURRENT FILING DATE: US/60/453,050
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 8028
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-050-8038

Query Match
Best Local Similarity 71.5%; Score 1831.5; DB 7; Length 370;
Matches 369; Conservative 0; Mismatches 1; Indels 125; Gaps

OY 1 MPPTGSPFFLLILLTLVTVVTSAGSASTPGCKETSATROSSVPSSTKNAVMSSTSV 60
Db 1 MPPTGSPFFLLILLTLVTVVTSAGSASTPGCKETSATROSSVPSSTKNAVMSSTSV 60
OY 61 LSSHPSGSGSSTTQODVTLAPATERPASGAATWGOVTVSVPTPALGSTTPPAHDVTS 120
Db 61 LSSHPSGSGSSTTQODVTLAPATERPASGAATWGOVTVSVPTPALGSTTPPAHDVTS 120
OY 121 APDNKPAVGTAPPAHVGTAPADTPRPPGSTAPPAHVGTAPADTPRPPGSTAPPAHVGT 180
Db 121 APDNKPAVGTAPPAHVGTAPADTPRPPGSTAPPAHVGTAPADTPRPPGSTAPPAHVGT 180
OY 181 APDNKPAVGTAPPAHVGTAPADTPRPPGSTAPPAHVGTAPADTPRPPGSTAPPAHVGT 240
Db 181 APDNKPAVGTAPPAHVGTAPADTPRPPGSTAPPAHVGTAPADTPRPPGSTAPPAHVGT 240
OY 241 PPTTAASTKTDASTTHNSVTPPLTSSNHSSTQSTGVSFFLSPIINLOPMSILED 300
Db 241 PPTTAASTKTDASTTHNSVTPPLTSSNHSSTQSTGVSFFLSPIINLOPMSILED 300
OY 221 PPTTAASTKTDASTTHNSVTPPLTSSNHSSTQSTGVSFFLSPIINLOPMSILED 280
Db 221 PPTTAASTKTDASTTHNSVTPPLTSSNHSSTQSTGVSFFLSPIINLOPMSILED 280
OY 301 PSTDYDELORDISEMFLQIKKGGSLGSLNKKRPSSVVQGLAPREGTIVHDVETQ 360
Db 281 PSTDYDELORDISEM----- 296
OY 361 PNOYKTBASRYVLTISDVSVHPFPFSAAGAGVPCMGIALVLVYVALATVYLLA 420
Db 297 ----- 296
OY 421 LAVQCGRRKRVYQGLDIPARPTVHNSKYPCTHGHGVNPPSTPRAPRYKVASGSGSS 480
Db 297 -AVCGRRKRVYQGLDIPARPTVHNSKYPCTHGHGVNPPSTPRAPRYKVASGSGSS 355
OY 481 LSTYTPAAVTAASNL 495
Db 356 LSTYTPAAVTAASNL 370

```

Dd	61	LSHSPGSGSTTQGDVUTLPAPEPAGSAAITWQDVTSPVTRPALSGTTPPAHDVTS	120
Qy	121	ABDNKPAAGSTAPPAAGVTSAPDTRPPGSGTAPAAHGVTSAPDTRPAPOSTPAPHGVTS	180
Dd	121	ABDNKPAAGSTAPP-----AHGVTSAPDTRPAPOSTPAPHGVTS	160
Qy	181	ABDNKPALGSTAPPVNNTVSAGSAGSAGSASTLVNHNSTARRATTPPAKSKTSPGISHSKD	240
Dd	161	ABDNKPALGSTAPPVNNTVSAGSAGSAGSASTLVNHNSTARRATTPPAKSKTSPGISHSKD	220
Qy	241	PTTTLASHSTRKASSTHSHSVPPLTGSNNHSTSPLSTGVSPFFPLSPFHLSNTOFNSSLED	300
Dd	221	PTTTLASHSTRKASSTHSHSVPPLTGSNNHSTSPLSTGVSPFFPLSPFHLSNTOFNSSLED	280
Qy	301	PSTDVYGELORDISEMFFQIYKQGGFLGSLNKKFPGSVVOLTLAFREGTINVDHYEQ	360
Dd	281	PSTDVYGELORDISEM-----	296
Qy	361	FMQYKTEAARNYMLTISDVSVSHVPPPSAOSGAGVPGMGIALVLVLCVLAVALVILA	420
Dd	297	-----	296
Qy	421	LAVCCCKRKNYGGLODIFPARDITYHPMSSEYPTVYHTRGVPPSSTDRSPYEKVSACNGGSS	480
Dd	297	-VYCCCKRKNYGGLODIFPARDITYHPMSSEYPTVYHTRGVPPSSTDRSPYEKVSACNGGSS	355
Qy	481	LSYTNPAVAATSAINTL495	
Dd	356	LSYTNPAVAATSAINTL370	

Search completed: May 1, 2003, 23:21:01
Job time : 110.089 secs

Thu May 6 16:14:02 2003

us-09-658-621b-2.rapm

Page 1

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model1

Run on: May 1, 2003, 23:10:31 ; Search time 273.036 Seconds
(without alignments)
1168.869 Million cell updates/sec

Title: US-09-658-621b-2

Sequence: 1 MPTGTQSPFPLILTLTVLTV.....NGGSLSYTNPAVAATSNL 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 64473110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Parents AA Main:
1: /cgn2_6/prodata/1/paa/PCTUS_COMB.pep.*
2: /cgn2_6/prodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/prodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/prodata/1/paa/US08_COMB.pep.*
5: /cgn2_6/prodata/1/paa/US081_COMB.pep.*
6: /cgn2_6/prodata/1/paa/US082_COMB.pep.*
7: /cgn2_6/prodata/1/paa/US083_COMB.pep.*
8: /cgn2_6/prodata/1/paa/US084_COMB.pep.*
9: /cgn2_6/prodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/prodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/prodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/prodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/prodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/prodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/prodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/prodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/prodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/prodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/prodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/prodata/1/paa/US096_COMB.pep.*
21: /cgn2_6/prodata/1/paa/US097_COMB.pep.*
22: /cgn2_6/prodata/1/paa/US098_COMB.pep.*
23: /cgn2_6/prodata/1/paa/US099_COMB.pep.*
24: /cgn2_6/prodata/1/paa/US100_COMB.pep.*
25: /cgn2_6/prodata/1/paa/US101_COMB.pep.*
26: /cgn2_6/prodata/1/paa/US102_COMB.pep.*
27: /cgn2_6/prodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2562	100.0	495	20	US-09-658-621-2
2	2516	98.2	515	1	PCT-US02-07826-212
3	2516	98.2	515	24	US-10-097-340-212
4	2516	98.2	515	25	US-10-171-311-156
5	2413	94.2	475	1	PCT-US01-30151-3
6	2413	94.2	475	20	US-09-658-621-67

7	2282.5	89.1	508	17	US-09-166-670-20	Sequence 20, Appl
8	2282.5	89.1	508	24	US-10-057-136-20	Sequence 20, Appl
9	2130	83.1	1255	1	PCT-US02-19669-311	Sequence 311, Appl
10	2130	83.1	1255	23	US-09-996-069-10	Sequence 10, Appl
11	2130	83.1	1255	25	US-10-171-311-158	Sequence 158, Appl
12	2130	83.1	1255	25	US-10-177-293-311	Sequence 311, Appl
13	1640	64.0	321	1	PCT-US00-05882-861	Sequence 861, Appl
14	1640	64.0	321	23	US-09-925-301-861	Sequence 861, Appl
15	1273	49.7	222	26	US-10-221-279-12252	Sequence 6112, Ap
16	1202.5	46.9	630	26	US-10-207-655-69	Sequence 12252, A
17	967.5	37.8	230	27	US-09-389-987-356	Sequence 356, App
18	967.5	37.8	230	27	US-09-412-418-356	Sequence 356, App
19	828	32.3	168	27	US-09-196-710-6165	Sequence 6165, App
20	828	32.3	168	27	US-09-196-718-6112	Sequence 6112, Ap
21	649	25.3	129	5	US-08-135-957-3	Sequence 3, Appl1
22	649	25.3	129	5	US-08-243-685-3	Sequence 3, Appl1
23	649	25.3	129	6	US-08-287-538-3	Sequence 3, Appl1
24	484	18.9	121	27	US-09-965-131-6	Sequence 3104, Ap
25	465	18.1	100	23	US-09-593-870A-12	Sequence 12, Appl
26	396	15.5	72	1	PCT-US01-49923-1	Sequence 1, Appl1
27	353	13.8	71	19	US-09-593-870A-12	Sequence 12, Appl
28	339	13.2	177	20	US-09-646-028-49	Sequence 49, Appl
29	334	13.0	172	20	US-09-646-028-49	Sequence 49, Appl
30	265	10.3	50	1	PCT-US01-49923-5	Sequence 5, Appl1
31	252	9.8	43	1	PCT-US01-49923-5	Sequence 5, Appl1
32	251	9.8	50	15	US-09-194-799C-12	Sequence 12, Appl
33	251	9.8	51	1	PCT-US97-04493-6	Sequence 6, Appl1
34	251	9.8	51	15	US-09-194-799C-6	Sequence 6, Appl1
35	251	9.8	51	15	US-09-194-799C-6	Sequence 6, Appl1
36	248	9.7	1953	27	US-09-185-361-566	Sequence 566, App
37	248	9.7	2947	27	US-09-185-361-566	Sequence 566, App
38	248	9.7	2947	27	US-09-185-361-566	Sequence 566, App
39	243	9.5	377	1	PCT-US01-49923-5	Sequence 5, Appl1
40	243	9.5	377	1	PCT-US01-49923-5	Sequence 5, Appl1
41	240	9.4	373	1	PCT-US00-32990-7	Sequence 7, Appl1
42	240	9.4	377	25	US-10-149-819-7	Sequence 7, Appl1
43	240	9.4	377	25	US-10-149-819-7	Sequence 7, Appl1
44	240	9.4	377	25	US-09-791-537-128709	Sequence 128709, Appl
45	234	9.1	753	27	US-09-185-568-351	Sequence 351, App

ALIGNMENTS

RESULT 1
US-09-658-621-2
/ Sequence 2, Application US/09658621
/ GENERAL INFORMATION:
/ APPLICANT: Taylor-Papadimitriou, Joyce
/ APPLICANT: Heukamp, Lukas Carl
/ APPLICANT: Offringa, Rienk
/ APPLICANT: Melief, Cornelis Johanna Maria
/ APPLICANT: Thomas, Mireille
/ TITLE OF INVENTION: MGC-1 derived peptides
/ FILE REFERENCE: 029395-017
/ CURRENT APPLICATION NUMBER: US/09/658, 621
/ PRIOR APPLICATION NUMBER: US 60/187, 215
/ PRIOR FILING DATE: 2000-03-03
/ PRIOR APPLICATION NUMBER: GB 9921242.5
/ PRIOR FILING DATE: 1999-09-08
/ PRIOR APPLICATION NUMBER: EP 99 40 2237.4
/ PRIOR FILING DATE: 1999-09-10
/ NUMBER OF SEQ IDS NOS: 67
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 495
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-658-621-2
Query Match 100.0% ; Score 2562; DB 20; Length 495;

Best Local Similarity 100.0%; Pred. No. 1,6e-160;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPTGQSPFFLLLLTLVTVTGSGHASTPGGKETSATGRSSVSSSTKXNAVMTSSV 60
Db 1 MPTGQSPFFLLLLTLVTVTGSGHASTPGGKETSATGRSSVSSSTKXNAVMTSSV 60
QY 61 LSSHPGSGSSTTQGDVTLAPATEPASGAATWGDVTSVPTRPALGSTTPPAHDVTS 120
Db 61 LSSHPGSGSSTTQGDVTLAPATEPASGAATWGDVTSVPTRPALGSTTPPAHDVTS 120
QY 121 APDNKPAAGTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 180
Db 121 APDNKPAAGTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 180
QY 181 APDNKPAAGTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 240
Db 181 APDNKPAAGTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 240
QY 241 TPTTLASHTKTDASTHSTHSTVPLTSSNHSJTSPLSTGVSPFLSHISNLQFNSSLED 300
Db 241 TPTTLASHTKTDASTHSTHSTVPLTSSNHSJTSPLSTGVSPFLSHISNLQFNSSLED 300
QY 301 PSTDYQELQDISEMFLQIYKQGGFLGSLNKKERGSVVQLTLAREGTTNVHDVETQ 360
Db 301 PSTDYQELQDISEMFLQIYKQGGFLGSLNKKERGSVVQLTLAREGTTNVHDVETQ 360
QY 361 PNOYKTEASRYNLITSDVSVSHPFPSSAQSAGVPGMGIALLVLCVVALAVYLIA 420
Db 361 PNOYKTEASRYNLITSDVSVSHPFPSSAQSAGVPGMGIALLVLCVVALAVYLIA 420
QY 421 LAVCCRRKNKGQLDIFPARDTYHMESEYPTVTHGRVYPPSSSTDSPEYKVSAGNASS 480
Db 421 LAVCCRRKNKGQLDIFPARDTYHMESEYPTVTHGRVYPPSSSTDSPEYKVSAGNASS 480
QY 481 LSTNPAAVATSANL 495
Db 481 LSTNPAAVATSANL 495

RESULT 2
PCT-US02-07826-212
Sequence 212, Application PC/TUS0207826
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc. et al.
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
FILE REFERENCE: MRI-030PC
CURRENT APPLICATION NUMBER: PCT/US02/07826
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 212
LENGTH: 515
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-07826-212

Query Match 98.2%; Score 2516; DB 1; Length 515;

Best Local Similarity 95.3%; Pred. No. 1,9e-157;
Matches 492; Conservative 0; Mismatches 3; Indels 20; Gaps 1;
QY 1 MPTGQSPFFLLLLTLVTVTGSGHASTPGGKETSATGRSSVSSSTKXNAVMTSSV 60
Db 1 MPTGQSPFFLLLLTLVTVTGSGHASTPGGKETSATGRSSVSSSTKXNAVMTSSV 60
QY 61 LSSHPGSGSSTTQGDVTLAPATEPASGAATWGDVTSVPTRPALGSTTPPAHDVTS 120
Db 61 LSSHPGSGSSTTQGDVTLAPATEPASGAATWGDVTSVPTRPALGSTTPPAHDVTS 120
QY 121 APDNKPAAGTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 160
Db 121 APDNKPAAGTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 180
QY 161 APDNKPAAGTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 220
Db 161 APDNKPAAGTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 240
QY 221 ATTPPASKSTPSSISHSHTDPTTLASHTKTDASTHSTHSTVPLTSSNHSJTSPLSTGV 280
Db 221 ATTPPASKSTPSSISHSHTDPTTLASHTKTDASTHSTHSTVPLTSSNHSJTSPLSTGV 300
QY 281 SFFFLSHISNLQFNSSLEDSDYQELQDISEMFLQIYKQGGFLGSLNKKERGSVV 340
Db 281 SFFFLSHISNLQFNSSLEDSDYQELQDISEMFLQIYKQGGFLGSLNKKERGSVV 360
QY 341 VOLTLAREGTTNVHDVETQFNQYKTEASRYNLITSDVSVSHPFPSSAQSAGVPGMG 400
Db 341 VOLTLAREGTTNVHDVETQFNQYKTEASRYNLITSDVSVSHPFPSSAQSAGVPGMG 420
QY 401 IALLVLCVVALAVYLIALAVCCRRKNKGQLDIFPARDTYHMESEYPTVTHGRVY 460
Db 401 IALLVLCVVALAVYLIALAVCCRRKNKGQLDIFPARDTYHMESEYPTVTHGRVY 480
QY 461 PSTDRSPYKVSAGNASSISTNPAAVATSANL 495
Db 461 PSTDRSPYKVSAGNASSISTNPAAVATSANL 515

RESULT 3
US-10-097-340-212
Sequence 212, Application US/10097340
GENERAL INFORMATION:
APPLICANT: John MONAHAN
APPLICANT: Manjula GANNAPARAPU
APPLICANT: Sebastian HOERSCH
APPLICANT: Shubhangi KAMATKAR
APPLICANT: Steve G. KOVATS
APPLICANT: Rachel B. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Peter VEIBY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumei ZHAO
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
FILE REFERENCE: MRI-030
CURRENT APPLICATION NUMBER: US/10/097,340
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26

Thu May 3 16:14:02 2003

us-09-658-621b-2.rapm

Page 3

```
/ PRIOR APPLICATION NUMBER: 60/311,732
/ PRIOR FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: 60/325,102
/ PRIOR FILING DATE: 2001-09-26
/ PRIOR APPLICATION NUMBER: 60/323,580
/ PRIOR FILING DATE: 2001-09-19
/ NUMBER OF SEQ ID NOS: 363
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 212
/ LENGTH: 515
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-097-340-212
```

```
Query Match 98.2%; Score 2516; DB 24; Length 515;
Best Local Similarity 95.5%; Pred. No. 1,9e-157;
Matches 492; Conservative 0; Mismatches 3; Indels 20; Gaps 1;
```

```
OY 1 MTGCTGSPFFLLLLTLVTVTSSGSHASTPGGKETSATQRSVSPSTKNAVSMTSV 60
DB 1 MTGCTGSPFFLLLLTLVTVTSSGSHASTPGGKETSATQRSVSPSTKNAVSMTSV 60
OY 61 LSSHSFGSSSTTGGODVTLAPATEPAGSAAATGQDVTSPVTRPALGSTPAPADVTS 120
DB 61 LSSHSFGSSSTTGGODVTLAPATEPAGSAAATGQDVTSPVTRPALGSTPAPADVTS 120
OY 121 APNKPAA-----PGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTS 160
DB 121 APNKPAPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTS 180
OY 161 APDTRPAPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTS 220
DB 161 APDTRPAPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTS 240
OY 221 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNSTSPOLSTGV 280
DB 221 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNSTSPOLSTGV 300
OY 281 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISSEMFQIYKGGFGLSNIKFRGSAV 340
DB 281 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISSEMFQIYKGGFGLSNIKFRGSAV 360
OY 341 VOLTAFREGTINHDVETQFNQYKTEAASRYNLTISDVSVDVPPFSAQSGAGVPGWG 400
DB 341 VOLTAFREGTINHDVETQFNQYKTEAASRYNLTISDVSVDVPPFSAQSGAGVPGWG 420
OY 401 IALLVVCVLAALVYLLALAVCCCRKNYGGDIFPAPDITYHMSYPTHTGRVYP 460
DB 401 IALLVVCVLAALVYLLALAVCCCRKNYGGDIFPAPDITYHMSYPTHTGRVYP 480
OY 461 PSSTDRSPYKVSAGNGSSLSYTNPAVAATSANL 495
DB 461 PSSTDRSPYKVSAGNGSSLSYTNPAVAATSANL 515
```

```
RESULT 4
US-10-171-311-156
/ Sequence 156, Application US/1017311
/ GENERAL INFORMATION:
/ APPLICANT: Schlegel, Robert
/ APPLICANT: Chen, Yan
/ APPLICANT: Zhao, Xumei
/ APPLICANT: Monahan, John
/ APPLICANT: Glatz, Karen
/ APPLICANT: Gannavarapu, Manjula
/ APPLICANT: Hoersb, Sebastian
/ TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
/ TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
/ FILE REFERENCE: MRI-035
/ CURRENT APPLICATION NUMBER: US/10/171,311
/ CURRENT FILING DATE: 2002-06-12
```

```
/ PRIOR APPLICATION NUMBER: US 60/296,159
/ PRIOR FILING DATE: 2001-06-13
/ PRIOR APPLICATION NUMBER: US 60/296,155
/ PRIOR FILING DATE: 2001-06-13
/ PRIOR APPLICATION NUMBER: US 60/335,936
/ PRIOR FILING DATE: 2001-11-14
/ NUMBER OF SEQ ID NOS: 238
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 156
/ LENGTH: 515
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-171-311-156
```

```
Query Match 98.2%; Score 2516; DB 25; Length 515;
Best Local Similarity 95.5%; Pred. No. 1,9e-157;
Matches 492; Conservative 0; Mismatches 3; Indels 20; Gaps 1;
```

```
OY 1 MTGCTGSPFFLLLLTLVTVTSSGSHASTPGGKETSATQRSVSPSTKNAVSMTSV 60
DB 1 MTGCTGSPFFLLLLTLVTVTSSGSHASTPGGKETSATQRSVSPSTKNAVSMTSV 60
OY 61 LSSHSFGSSSTTGGODVTLAPATEPAGSAAATGQDVTSPVTRPALGSTPAPADVTS 120
DB 61 LSSHSFGSSSTTGGODVTLAPATEPAGSAAATGQDVTSPVTRPALGSTPAPADVTS 120
OY 121 APNKPAA-----PGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTS 160
DB 121 APNKPAPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTS 180
OY 161 APDTRPAPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTS 220
DB 161 APDTRPAPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTS 240
OY 221 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNSTSPOLSTGV 280
DB 221 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNSTSPOLSTGV 300
OY 281 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISSEMFQIYKGGFGLSNIKFRGSAV 340
DB 281 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISSEMFQIYKGGFGLSNIKFRGSAV 360
OY 341 VOLTAFREGTINHDVETQFNQYKTEAASRYNLTISDVSVDVPPFSAQSGAGVPGWG 400
DB 341 VOLTAFREGTINHDVETQFNQYKTEAASRYNLTISDVSVDVPPFSAQSGAGVPGWG 420
OY 401 IALLVVCVLAALVYLLALAVCCCRKNYGGDIFPAPDITYHMSYPTHTGRVYP 460
DB 401 IALLVVCVLAALVYLLALAVCCCRKNYGGDIFPAPDITYHMSYPTHTGRVYP 480
OY 461 PSSTDRSPYKVSAGNGSSLSYTNPAVAATSANL 495
DB 461 PSSTDRSPYKVSAGNGSSLSYTNPAVAATSANL 515
```

```
RESULT 5
PCT-US01-30151-3
/ Sequence 3, Application PC/TUS0130151
/ GENERAL INFORMATION:
/ APPLICANT: GenSisance Pharmaceutical, Inc.
/ APPLICANT: Koshiy, Beena
/ APPLICANT: Koshiy, Beena
/ TITLE OF INVENTION: HAPLOTYPES OF THE MUC1 GENE
/ FILE REFERENCE: MUC1 MMH-1408PCT
/ CURRENT APPLICATION NUMBER: PCT/US01/30151
/ PRIOR FILING DATE: 2001-09-25
/ PRIOR APPLICATION NUMBER: 60/236,113
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 3
/ LENGTH: 475
/ TYPE: PRT
```

Thu May 8 16:14:02 2003

us-09-658-621b-2.ram

Page 4

```

/ ORGANISM: Homo sapiens
PCT-US01-30151-3

Query Match      94.2%; Score 2413; DB 1; Length 475;
Best Local Similarity 95.6%; Pred. No. 1,1e-150;
Matches 473; Conservative 0; Mismatches 2; Indels 20; Gaps 1;

QY 1 MTPGOSPFLLILITLVLTVTGSGHASTPGGEKETSATORSVSSSTEKNAISMSSY 60
DB 1 MTPGOSPFLLILITLVLTVTGSGHASTPGGEKETSATORSVSSSTEKNAISMSSY 60
QY 61 LSSSPSGSGSTTQGGDYTLAPATERPAGSAATWGCQDVTSVPTRPALGSTTPPAHNVTS 120
DB 61 LSSSPSGSGSTTQGGDYTLAPATERPAGSAATWGCQDVTSVPTRPALGSTTPPAHNVTS 120
QY 121 APDKRPAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
DB 121 APDKRPAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
QY 181 APDKRPAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
DB 181 APDKRPAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
QY 241 TPTTLASHTKTDASTHSTVPLTSSNHSTSPOLSTGVSEFFLSFHI.SMLQPNSSLED 300
DB 241 TPTTLASHTKTDASTHSTVPLTSSNHSTSPOLSTGVSEFFLSFHI.SMLQPNSSLED 300
QY 301 PSTDYOELOKDISMFLQIKQGGFLGSLNKRPGSVVVOULTLAFREGTINVDVETO 360
DB 301 PSTDYOELOKDISMFLQIKQGGFLGSLNKRPGSVVVOULTLAFREGTINVDVETO 360
QY 361 FNOKTEAARVNLITISVSVSHVPPFSAQSGAGVPGWGIALLVYCVLAVALIYVLLA 420
DB 361 FNOKTEAARVNLITISVSVSHVPPFSAQSGAGVPGWGIALLVYCVLAVALIYVLLA 420
QY 421 LAVOCCKRNKYGQDLPFARDTYHPMSEYFTHYHGRVYPPSSSTDRSPYEKVSANGCGSS 480
DB 421 LAVOCCKRNKYGQDLPFARDTYHPMSEYFTHYHGRVYPPSSSTDRSPYEKVSANGCGSS 480
QY 481 LSYNPAVAATSANL 495
DB 481 LSYNPAVAATSANL 495
DB 461 LSYNPAVAATSANL 475

RESULT 6
US-09-658-621-67
/ Sequence 67, Application US/09658621
/ GENERAL INFORMATION:
/ APPLICANT: Taylor-Papadimitriou, Joyce
/ APPLICANT: Heukamp, Lukas Carl
/ APPLICANT: Oettinger, Rieck
/ APPLICANT: Meisler, Cornelia Johanna Maria
/ APPLICANT: Acres, Bruce
/ APPLICANT: Thomas, Mireille
/ TITLE OF INVENTION: KCC-1 derived peptides
/ FILE REFERENCE: 023995-017
/ CURRENT APPLICATION NUMBER: US/09/658, 621
/ PRIOR FILING DATE: 2000-09-08
/ PRIOR APPLICATION NUMBER: US 60/187,215
/ PRIOR FILING DATE: 2000-03-03
/ PRIOR APPLICATION NUMBER: 9921242.5
/ PRIOR FILING DATE: 1999-09-08
/ PRIOR APPLICATION NUMBER: EP 99 40 2237.4
/ NUMBER OF SEQ ID NOS: 10
/ SOFTWARE: Patent Ver. 2.1
/ SEQ ID NO 67
/ LENGTH: 475
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-658-621-67

Query Match      94.2%; Score 2413; DB 20; Length 475;
```

```

Best Local Similarity 95.6%; Pred. No. 1,1e-150;
Matches 473; Conservative 0; Mismatches 2; Indels 20; Gaps 1;

QY 1 MTPGOSPFLLILITLVLTVTGSGHASTPGGEKETSATORSVSSSTEKNAISMSSY 60
DB 1 MTPGOSPFLLILITLVLTVTGSGHASTPGGEKETSATORSVSSSTEKNAISMSSY 60
QY 61 LSSSPSGSGSTTQGGDYTLAPATERPAGSAATWGCQDVTSVPTRPALGSTTPPAHNVTS 120
DB 61 LSSSPSGSGSTTQGGDYTLAPATERPAGSAATWGCQDVTSVPTRPALGSTTPPAHNVTS 120
QY 121 APDKRPAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
DB 121 APDKRPAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
QY 181 APDKRPAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
DB 181 APDKRPAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
QY 241 TPTTLASHTKTDASTHSTVPLTSSNHSTSPOLSTGVSEFFLSFHI.SMLQPNSSLED 300
DB 241 TPTTLASHTKTDASTHSTVPLTSSNHSTSPOLSTGVSEFFLSFHI.SMLQPNSSLED 300
QY 301 PSTDYOELOKDISMFLQIKQGGFLGSLNKRPGSVVVOULTLAFREGTINVDVETO 360
DB 301 PSTDYOELOKDISMFLQIKQGGFLGSLNKRPGSVVVOULTLAFREGTINVDVETO 360
QY 361 FNOKTEAARVNLITISVSVSHVPPFSAQSGAGVPGWGIALLVYCVLAVALIYVLLA 420
DB 361 FNOKTEAARVNLITISVSVSHVPPFSAQSGAGVPGWGIALLVYCVLAVALIYVLLA 420
QY 421 LAVOCCKRNKYGQDLPFARDTYHPMSEYFTHYHGRVYPPSSSTDRSPYEKVSANGCGSS 480
DB 421 LAVOCCKRNKYGQDLPFARDTYHPMSEYFTHYHGRVYPPSSSTDRSPYEKVSANGCGSS 480
QY 481 LSYNPAVAATSANL 495
DB 481 LSYNPAVAATSANL 495
DB 461 LSYNPAVAATSANL 475

RESULT 7
US-09-366-670-20
/ Sequence 20, Application US/09366670
/ GENERAL INFORMATION:
/ APPLICANT: SCHLOM, JEFFREY
/ APPLICANT: KANTOR, JUDITH
/ APPLICANT: KUPE, DONALD
/ APPLICANT: PANICALLY, DENNIS
/ TITLE OF INVENTION: RECOMBINANT FOX VIRUS FOR IMMUNIZATION AGAINST MUC1
/ FILE REFERENCE: 700953/47113C
/ CURRENT APPLICATION NUMBER: US/09/366, 670
/ PRIOR FILING DATE: 1999-08-03
/ EARLIER APPLICATION NUMBER: PCT/US98/03693
/ EARLIER FILING DATE: 1998-02-24
/ EARLIER APPLICATION NUMBER: 60/038, 253
/ NUMBER OF SEQ ID NOS: 20
/ SOFTWARE: Patent Ver. 2.1
/ SEQ ID NO 20
/ LENGTH: 508
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-366-670-20

Query Match      89.1%; Score 2282.5; DB 17; Length 508;
Best Local Similarity 89.2%; Pred. No. 4.9e-142;
Matches 455; Conservative 6; Mismatches 32; Indels 17; Gaps 3;

QY 1 MTPGOSPFLLILITLVLTVTGSGHASTPGGEKETSATORSVSSSTEKNAISMSSY 60
DB 1 MTPGOSPFLLILITLVLTVTGSGHASTPGGEKETSATORSVSSSTEKNAISMSSY 60
```

Thu May 8 16:14:02 2003

us-09-658-621b-2.ram

Page 5

Qy	52	NISMTSTVLSLSSHPSSGSGSTTGQGDVTLA	PA	TE	PA	SS	GA	MT	MG	OD	VT	SV	PT	PA	LG	ST	11.11	
Qy	61	MAVSTTS	-	LL	NS	CV	RGS	T	A	P	A	H	CV	T	S	A	PD	11.88
Db	112	TPADHNVSAADNKP	AG	ST	A	P	AG	T	S	A	P	D	R	-	-	-	16.55	
Qy	119	APAHGVISADPTR	PA	PG	ST	A	P	AG	T	S	A	P	D	R	A	GS	17.87	
Db	166	PARGSTAPPAHCV	S	A	P	D	R	P	A	LG	S	T	A	P	A	P	22.25	
Qy	179	PARGSTAPPAHCV	S	A	P	D	R	P	A	LG	S	T	A	P	A	P	23.88	
Db	226	AKSGTPBSIBH	HD	PT	L	A	SH	S	K	T	D	A	S	T	H	S	26.85	
Qy	239	AKSGTSPBSIBH	HD	PT	L	A	SH	S	K	T	D	A	S	T	H	S	29.68	
Db	286	SPTLSNLP	GN	S	L	E	D	P	S	T	D	Y	O	E	L	O	34.54	
Qy	299	SPTLSNLP	GN	S	L	E	D	P	S	T	D	Y	O	E	L	O	35.88	
Db	346	AFREGTIN	H	D	V	E	T	O	E	N	K	I	E	P	A	S	40.55	
Qy	359	AFREGTIN	H	D	V	E	T	O	E	N	K	I	E	P	A	S	41.88	
Db	406	LVCVVALA	I	V	A	L	A	V	C	O	C	R	K	Y	G	O	44.55	
Qy	419	LVCVVALA	I	V	A	L	A	V	C	O	C	R	K	Y	G	O	47.88	
Db	466	RSEYEK	V	S	A	G	S	S	L	E	Y	N	P	V	A	S	49.55	
Qy	479	RSEYEK	V	S	A	G	S	S	L	E	Y	N	P	V	A	S	50.88	

```

US-10-057-136-20
; Sequence 20, Application us/10057136
; GENERAL INFORMATION:
; APPLICANT: SCHLOW, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUFE, DONALD
; APPLICANT: PANICALI, DENNIS
; APPLICANT: GRITZ, LINDA
; TITLE OF INVENTION: RECOMBINANT FOX VIRUS FOR IMMUNIZATION AGAINST MUC1
; FILE REFERENCE: 7009593,4/7113C
; CURRENT APPLICATION NUMBER: US/10/057,136
; CURRENT PRIORITY NUMBER: 06/0122
; PRIOR FILING DATE: 1989-08-03,366,670
; PRIOR APPLICATION NUMBER: PCT/US98/036993
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/038,253
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-136-20

Query Match: 89.1%; Score 2292.5; DB 24; Length 508;
Desc: Local Similarity 89.2%; E-val: 4.9e-14; Ids: 17; Gaps
Matches 45; Conservative 6; Mismatches 37;

CY 1 MFGGSGSPPLLLLTLYL-----IVYSGHMSSTPGGKERTASTORSVSBSSTK 51
DB 1 MFGGSGSPPLLLLTLYLATTAPKPAIVTGGSGHASTPGGKERTASTORSVSBSSTK 60

CY 52 NAVSMSTSVLASHSPGGSGSTTCOGDVTLLAPATPSAGSAATWOGDVTSPVTPRPALGST 11
DB 61 NAVSMSTSVLASHSPGGSGSTTCOGDVTLLAPATPSAGSAATWOGDVTSPVTPRPALGST 118

```

Qy	112	ATCAIDVTSAPDKRPAAGSTAPPAHAGVTSAPDR-----	-PPPGSTAPPAHAGVTSAPDR	168
Db	119	APPAHAGVTSAPDRPAPAGSTAPPAHAGVTSAPDRRAPAGSTAPAPSTAPPAHAGVTSAPDR		178
Qy	166	PAPGSTAPPAHAGVTSAPDRPALGSTAPPAHNTSASGSAGSASTLVHNGTSASATTP		222
Db	179	PAPGSTAPPAHAGVTSAPDRPALGSTAPPAHNTSASGSAGSASTLVHNGTSASATTP		233
Qy	226	AKSTPPSIPSHSDPPTLLASHTKDASTHSTVPELTTSSNHSSTPGLSTGVSPFEL		285
Db	239	AKSTPPSIPSHSDPPTLLASHTKDASTHSTVPELTTSSNHSSTPGLSTGVSPFEL		296
Dy	286	SHHSISNIPFNSSIEDPTDYQSIORDISMFQTIKKGGELGSIKTPRGSVVOVTL		345
Qy	299	SHHSISNIPFNSSIEDPTDYQSIORDISMFQTIKKGGELGSIKTPRGSVVOVTL		355
Dy	346	APRSTINHDVETPONQKTEAASRNLTISVSGSVPPPSAOSGAGVSGMGIALLY		405
Db	359	APRSTINHDVETPONQKTEAASRNLTIPVSGSDVPPPSAOSGAGVSGMGIALLY		415
Qy	406	LVCULVALAIVYLIALAVOCRRKNYGQDLIFPARDTYHPMSSEVYTHHGKRVPPSSTD		465
Db	419	LVCULVALAIVYLIALAVOCRRKNYGQDLIFPARDTYHPMSSEVYTHHGKRVPPSSTD		475
Qy	466	RSPEPKVASNGGSSLSYTNPAVVAATSANL	495	
Db	479	RSPEPKVASNGGSSLSYTNPAVVAATSANL	508	

```

RESULT 9
PCT-US02-19669-311
: Sequence 311, Application PC/TUS0219669
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc. et al.
: TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
: TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
: FILE REFERENCE: MRI-038PC
: CURRENT APPLICATION NUMBER: PCT/US02/19669
: PCT FILING DATE: 2002-06-21
: PRIOR APPLICATION NUMBER: US 60/299,887
: PRIOR FILING DATE: 2001-06-21
: PRIOR APPLICATION NUMBER: US 60/301,572
: PRIOR FILING DATE: 2001-06-27
: PRIOR APPLICATION NUMBER: US 60/306,501
: PRIOR FILING DATE: 2001-07-18
: PRIOR APPLICATION NUMBER: US 60/325,002
: PRIOR FILING DATE: 2001-09-25
: PRIOR APPLICATION NUMBER: US 60/362,585
: PRIOR FILING DATE: 2002-03-05
: PRIOR APPLICATION NUMBER: US 60/xxx,xxx
: PRIOR FILING DATE: 2002-05-14
: NUMBER OF SEQ ID NOS: 506
: SOURCE: SEQ ID NO 311reSeq for Windows Version 4.0
: TYPE: PRT
: LENGTH: 1255
: ORGANISM: Homo sapiens
PCT-US02-19669-311

Query March      83.1%: Score 2130; DB 1; Length 1255;
Best Local Similarity 39.0%: Pred. No. 1,8e-131;
Matches 489; Conservative 1; Mismatches 5; Indels 760; Gaps 1;

Qy      1  MPTGQSPFLLILLTLVTYVTSQGHASTPGGKXKTSATQRSVPSSTEDKNAVMTSSV 60
Db      1  MPTGQSPFLLILLTLVTYVTSQGHASTPGGKXKTSATQRSVPSSTEDKNAVMTSSV 60

Qy      6  LSSHSPPSGSGSTTQGGQDVTLAPATEPAGSAAITVGGQDVTSVPVTPPALGSGTTPADVTS 120
Db      6  LSSHSPPSGSGSTTQGGQDVTLAPATEPAGSAAITVGGQDVTSVPVTPPALGSGTTPADVTS 120

Qy     121  APDVKKAPAGSTAPPAHGVTSAPDTRPPPSSTAPPAHGVTSAPDT----- 164
Db     121  APDVKKAPAGSTAPPAHGVTSAPDTRPPPSSTAPPAHGVTSAPDT----- 164

```

Db 121 APDKKPAAGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 180
Qy 165 ----- 164
Db 181 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 240
Qy 165 ----- 164
Db 241 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 300
Qy 165 ----- 164
Db 301 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 360
Qy 165 ----- 164
Db 361 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 420
Qy 165 ----- 164
Db 421 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 480
Qy 165 ----- 164
Db 481 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 540
Qy 165 ----- 164
Db 541 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 600
Qy 165 ----- 164
Db 601 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 660
Qy 165 ----- 164
Db 661 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 720
Qy 165 ----- 164
Db 721 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 780
Qy 165 ----- 164
Db 781 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 840
Qy 165 ----- 164
Db 841 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 900
Qy 165 ----- 164
Db 901 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 960
Qy 201 ASGASGASGASTLVHNGTSARATTPPASKSTPPSPSPSHSOTPTTLASHTKSTASSTHS 260
Db 961 ASGASGASGASTLVHNGTSARATTPPASKSTPPSPSPSHSOTPTTLASHTKSTASSTHS 1020
Qy 261 TVPPLTSSMNSTPOLSTGVFFFLSFHISNLQFNSSLEPSTDYQLOLQDISSEMFLOI 320
Db 1021 SVPLTSSMNSTPOLSTGVFFFLSFHISNLQFNSSLEPSTDYQLOLQDISSEMFLOI 1080
Qy 321 YKGGPFGLSNIFKRRPGSVVUOLTLAFREGTINWHDVETOFNOKYTEAASRYNLITSDVS 380
Db 1081 YKGGPFGLSNIFKRRPGSVVUOLTLAFREGTINWHDVETOFNOKYTEAASRYNLITSDVS 1140
Qy 381 VSHVPPFPFSAOSGAGVPGMGIALVLVLCVLVLAIVLILALANCCCRKXNGOLDIFPAR 440
Db 1141 VSDVPPFPFSAOSGAGVPGMGIALVLVLCVLVLAIVLILALANCCCRKXNGOLDIFPAR 1200
Qy 441 DTYHPMSRYPTHTHGRVVPSTDRSPYEKVSAGNGGSSLSTNPAAVAASNL 495

Db 1201 DTYHPMSRYPTHTHGRVVPSTDRSPYEKVSAGNGGSSLSTNPAAVAASNL 1255
RESULT 10
US-09-996-069-10
; Sequence 10, Application us/09996069
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia
; APPLICANT: Bamdad, R. Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBITORS
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996, 069
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 10
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-10
Query Match 83.1%; Score 2130; DB 23; Length 1255;
Best Local Similarity 39.0%; Pred. No. 1.8e-131;
Matches 489; Conservative 1; Mismatches 5; Indels 760; Gaps 1;
Qy 1 MTPGTQSPFLLLLTLVLTVTYTGSGHASTPGGEKETISATQSSVPSSTERNVAMTSV 60
Db 1 MTPGTQSPFLLLLTLVLTVTYTGSGHASTPGGEKETISATQSSVPSSTERNVAMTSV 60
Qy 61 LSHSPGSGSSTTQGDVTLAPATEPASGAATWGDVTSVPTRPALGSTTPPAHGVTS 120
Db 61 LSHSPGSGSSTTQGDVTLAPATEPASGAATWGDVTSVPTRPALGSTTPPAHGVTS 120
Qy 121 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 164
Db 121 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 180
Qy 165 ----- 164
Db 181 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 240
Qy 165 ----- 164
Db 241 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 300
Qy 165 ----- 164
Db 301 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 360
Qy 165 ----- 164
Db 361 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 420
Qy 165 ----- 164
Db 421 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 480
Qy 165 ----- 164
Db 481 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 540
Qy 165 ----- 164
Db 541 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 600
Qy 165 ----- 164
Db 601 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 660
Qy 165 ----- 164
Db 661 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 720

Qy 165 ----- 164
Db 721 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 780
Qy 165 ----- 164
Db 781 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 840
Qy 165 ----- 164
Db 841 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 900
Qy 165 ----- 200
Db 901 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 960
Qy 201 AAGSAGSAGSTLVHNGTSARATTPPASKSTPPSPSHHSDPTTLASHSTKTDASSTHHS 260
Db 961 AAGSAGSAGSTLVHNGTSARATTPPASKSTPPSPSHHSDPTTLASHSTKTDASSTHHS 1020
Qy 261 TVPPLTSNNHSTSPOLSTGVSPFLSPHISNLOFNSSLSDPTDYOELOQDISSEMFLOI 320
Db 1021 SVPLPITSNNHSTSPOLSTGVSPFLSPHISNLOFNSSLSDPTDYOELOQDISSEMFLOI 1080
Qy 321 YKGGFLSLNKKRPGSVVOLTARFGSTINMVDVETOPNOYKTEASRYNLIDVGS 360
Db 1081 YKGGFLSLNKKRPGSVVOLTARFGSTINMVDVETOPNOYKTEASRYNLIDVGS 1140
Qy 381 VSHVPPPSAAGSAGVPGWGLALVLCVLMALVYLIALAVQCCRRKRYGOLDIFPAR 440
Db 1141 VSHVPPPSAAGSAGVPGWGLALVLCVLMALVYLIALAVQCCRRKRYGOLDIFPAR 1200
Qy 441 DTYHMSGYPTTHHGRVVPSTDRSPRYEKVAGGSSSLSTYNPAVAATNANI 495
Db 1201 DTYHMSGYPTTHHGRVVPSTDRSPRYEKVAGGSSSLSTYNPAVAATNANI 1255

RESULT 11
US-10-171-311-158
; Sequence 158, Application US/10171311
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamakazi, Shubhangi
; APPLICANT: Glati, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoarish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 158
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-158
Query Match 83.1%; Score 2130; DB 25; Length 1255;
Best Local Similarity 39.0%; Pred. No. 1,8e-13;
Matches 489, Conservative 1; Mismatches 5; Indels 760; Gaps 1;

Qy 1 MTGGTGGFFLLILITVTYVGGSGASTBCKEKTASORSSVPSSTERNAMVMSSTSV 60
Db 1 MTGGTGGFFLLILITVTYVGGSGASTBCKEKTASORSSVPSSTERNAMVMSSTSV 60
Qy 61 LSHSFGSGSTTQGGQVTLAPATERASGAATWGGDVTSPVTPRPAIGSTSPADHVT 120
Db 61 LSHSFGSGSTTQGGQVTLAPATERASGAATWGGDVTSPVTPRPAIGSTSPADHVT 120
Qy 121 AADNKPRAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 164
Db 121 AADNKPRAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 180
Qy 165 ----- 164
Db 181 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 240
Qy 165 ----- 164
Db 241 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 300
Qy 165 ----- 164
Db 301 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 360
Qy 165 ----- 164
Db 361 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 420
Qy 165 ----- 164
Db 421 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 480
Qy 165 ----- 164
Db 481 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 540
Qy 165 ----- 164
Db 541 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 600
Qy 165 ----- 164
Db 601 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 660
Qy 165 ----- 164
Db 661 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 720
Qy 165 ----- 164
Db 721 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 780
Qy 165 ----- 164
Db 781 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 840
Qy 165 ----- 164
Db 841 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 900
Qy 165 ----- 200
Db 901 AADTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 960
Qy 201 AAGSAGSAGSTLVHNGTSARATTPPASKSTPPSPSHHSDPTTLASHSTKTDASSTHHS 260
Db 961 AAGSAGSAGSTLVHNGTSARATTPPASKSTPPSPSHHSDPTTLASHSTKTDASSTHHS 1020
Qy 261 TVPPLTSNNHSTSPOLSTGVSPFLSPHISNLOFNSSLSDPTDYOELOQDISSEMFLOI 320
Db 1021 SVPLPITSNNHSTSPOLSTGVSPFLSPHISNLOFNSSLSDPTDYOELOQDISSEMFLOI 1080

```
QY 321 YKGGFLGSLNIFKRPSSVVVQTLTAFBGGTINVDVETQPNQYKTEAASRYNLTISVS 380
DB 1081 YKGGFLGSLNIFKRPSSVVVQTLTAFBGGTINVDVETQPNQYKTEAASRYNLTISVS 1140
QY 361 VSNVPPFSSAGSAGVPGMGIALVLCVLAVALVYLLALAVCCCRKNYGGOLDIFPAR 440
DB 1141 VSDVPPFSSAGSAGVPGMGIALVLCVLAVALVYLLALAVCCCRKNYGGOLDIFPAR 1200
QY 441 DTHPMSSRYPTHTHGRVYVPPSSDTSRSEYENASGNGSSSLSTNPAVAATSANL 495
DB 1201 DTHPMSSRYPTHTHGRVYVPPSSDTSRSEYENASGNGSSSLSTNPAVAATSANL 1255

RESULT 12
Sequence 311, Application US/10177293
GENERAL INFORMATION:
APPLICANT: NPROLACTON, James
APPLICANT: Elliott, James
APPLICANT: Zhao, Xumei
APPLICANT: Gannavarapu, Manjula
APPLICANT: Kamatekar, Shubhangi
APPLICANT: Merrens, Maureen
APPLICANT: Meyer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Bae, J., Robert C.
APPLICANT: Horrobay, Gabriel N.
APPLICANT: Pustet, LaJos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegül
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
FILE REFERENCE: M81-018
FILE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
CURRENT APPLICATION NUMBER: US/10177,293
PRIOR FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 311
LENGTH: 1255
TYPE: PR1
ORGANISM: Homo sapiens
US-10-177-293-311

Query Match 83.1%; Score 2130; DB 25; Length 1255;
Best Local Similarity 39.0%; Pred. No. 1,8e-131;
Matches 489; Conservative 1; Mismatches 5; Indels 760; Gaps 1;
```

```
DB 121 APDNKPAAGSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 180
QY 165 ----- 164
DB 181 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 240
QY 165 ----- 164
DB 241 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 300
QY 165 ----- 164
DB 301 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 360
QY 165 ----- 164
DB 361 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 420
QY 165 ----- 164
DB 421 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 480
QY 165 ----- 164
DB 481 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 540
QY 165 ----- 164
DB 541 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 600
QY 165 ----- 164
DB 601 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 660
QY 165 ----- 164
DB 661 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 720
QY 165 ----- 164
DB 721 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 780
QY 165 ----- 164
DB 781 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 840
QY 165 ----- 164
DB 841 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 900
QY 165 ----- 164
DB 901 APDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTSAPDTRPAPOSTAPPAHGVTS 960
QY 201 ASGSASGASATVHNGTSABATTPPAKSTPSTPSHSDPTTLASHTDASSTHHS 260
DB 961 ASGSASGASATVHNGTSABATTPPAKSTPSTPSHSDPTTLASHTDASSTHHS 1020
QY 261 TVPPLTSSNHSSTPOLSTGVSPFLSFHISNLFNSGLSDPSDYKOLQDISMFLQI 320
DB 1021 SVPLTSSNHSSTPOLSTGVSPFLSFHISNLFNSGLSDPSDYKOLQDISMFLQI 1080
QY 321 YKGGFLGSLNIFKRPSSVVVQTLTAFBGGTINVDVETQPNQYKTEAASRYNLTISVS 380
DB 1081 YKGGFLGSLNIFKRPSSVVVQTLTAFBGGTINVDVETQPNQYKTEAASRYNLTISVS 1140
QY 381 VSNVPPFSSAGSAGVPGMGIALVLCVLAVALVYLLALAVCCCRKNYGGOLDIFPAR 440
DB 1141 VSDVPPFSSAGSAGVPGMGIALVLCVLAVALVYLLALAVCCCRKNYGGOLDIFPAR 1200
QY 441 DTHPMSSRYPTHTHGRVYVPPSSDTSRSEYENASGNGSSSLSTNPAVAATSANL 495
```


Thu May 8 16:14:02 2003

us-09-658-621b-2.ram

Page 9

```
Db 1201 DTHHMSSEPTTTHGRVYPPSSDTRSPYKYSAGNGSSLSSTYNPAVAASNL 1255
RESULT 13
PCT-US00-05882-861
Sequence 861, Application PC/TUS0005882
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben,
TITLE OF INVENTION: Human Cancer Associated Gene Sequences and Polypeptides
FILE REFERENCE: P106PCT
CURRENT APPLICATION NUMBER: PCT/US00/05882
CURRENT FILING DATE: 2000-03-08
EARLIER APPLICATION NUMBER: 60/124,270
EARLIER FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 861
LENGTH: 321
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (18)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US00-05882-861

Query Match 64.0%; Score 1640; DB 1; Length 321;
Best Local Similarity 99.1%; Pred. No. 7,6e-100;
Matches 318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Query 175 AGCVTSAPDNRPALGSTAPPVNNTVSAGSAGSASTLVHNGTSARATTTTPASKSTPSSI 234
Db 1 AGCVTSAPDNRPALGSTAPPVNNTVSAGSAGSASTLVHNGTSARATTTTPASKSTPSSI 60
Query 235 PSHHSDPTTTLASHSTKTDAASSTHSTVPPVLTSSNHSSTSPOLSTGVSEFFLSTHLSNQF 294
Db 61 PSHHSDPTTTLASHSTKTDAASSTHSTVPPVLTSSNHSSTSPOLSTGVSEFFLSTHLSNQF 120
Query 295 NSSLBPDSTDYQELQDISEMFLQIKYKGGFLGLSNIKFRPGSVVVQTLAFREGTINV 354
Db 121 NSSLBPDSTDYQELQDISEMFLQIKYKGGFLGLSNIKFRPGSVVVQTLAFREGTINV 180
Query 355 HDVETQFNQYKTEAASRYNLTISDVSVSHVPPPSAOSGAGVPGMGIALLVLCYVALA 414
Db 181 HDVETQFNQYKTEAASRYNLTISDVSVSHVPPPSAOSGAGVPGMGIALLVLCYVALA 240
Query 415 IYVLLALAVCCCRKXNYGQDIFPARDTYHPMSEYPTVTHGRVYPPSSDTRSPYKYS 474
Db 241 IYVLLALAVCCCRKXNYGQDIFPARDTYHPMSEYPTVTHGRVYPPSSDTRSPYKYS 300
Query 475 GNGSSLSSTYNPAVAASNL 495
Db 301 GNGSSLSSTYNPAVAASNL 321

RESULT 14
US-09-925-301-861
Sequence 861, Application US/09925301
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 861
LENGTH: 321
```

```
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (18)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-861

Query Match 64.0%; Score 1640; DB 23; Length 321;
Best Local Similarity 99.1%; Pred. No. 7,6e-100;
Matches 318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Query 175 AGCVTSAPDNRPALGSTAPPVNNTVSAGSAGSASTLVHNGTSARATTTTPASKSTPSSI 234
Db 1 AGCVTSAPDNRPALGSTAPPVNNTVSAGSAGSASTLVHNGTSARATTTTPASKSTPSSI 60
Query 235 PSHHSDPTTTLASHSTKTDAASSTHSTVPPVLTSSNHSSTSPOLSTGVSEFFLSTHLSNQF 294
Db 61 PSHHSDPTTTLASHSTKTDAASSTHSTVPPVLTSSNHSSTSPOLSTGVSEFFLSTHLSNQF 120
Query 295 NSSLBPDSTDYQELQDISEMFLQIKYKGGFLGLSNIKFRPGSVVVQTLAFREGTINV 354
Db 121 NSSLBPDSTDYQELQDISEMFLQIKYKGGFLGLSNIKFRPGSVVVQTLAFREGTINV 180
Query 355 HDVETQFNQYKTEAASRYNLTISDVSVSHVPPPSAOSGAGVPGMGIALLVLCYVALA 414
Db 181 HDVETQFNQYKTEAASRYNLTISDVSVSHVPPPSAOSGAGVPGMGIALLVLCYVALA 240
Query 415 IYVLLALAVCCCRKXNYGQDIFPARDTYHPMSEYPTVTHGRVYPPSSDTRSPYKYS 474
Db 241 IYVLLALAVCCCRKXNYGQDIFPARDTYHPMSEYPTVTHGRVYPPSSDTRSPYKYS 300
Query 475 GNGSSLSSTYNPAVAASNL 495
Db 301 GNGSSLSSTYNPAVAASNL 321

RESULT 15
US-10-221-279-12252
Sequence 12252, Application US/10221279
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 21272-046
CURRENT APPLICATION NUMBER: US/10/221,279
CURRENT FILING DATE: 2002-09-06
PRIOR APPLICATION NUMBER: 09/574,454
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: 09/519,705
PRIOR FILING DATE: 2000-03-07
NUMBER OF SEQ ID NOS: 12360
SOFTWARE: Cseton
SEQ ID NO 12252
LENGTH: 292
TYPE: PRT
ORGANISM: Homo sapiens
US-10-221-279-12252

Query Match 49.7%; Score 1273; DB 26; Length 292;
Best Local Similarity 54.9%; Pred. No. 1.2e-75;
Matches 272; Conservative 0; Mismatches 1; Indels 22; Gaps 1;

Query 1 MTGTGSPPLLLILVLYVVGSGHASTTGAKXTSATYORSSVPSSTKRNAMVMTSSV 60
Db 20 MTGTGSPPLLLILVLYVVGSGHASTTGAKXTSATYORSSVPSSTKRNAMVMTSSV 72
Query 61 LSHSPGSGSSTTGQDVTIAPTEPASGAATMGQDVTSVVTRPALSTTPPAHDVTS 120
Db 73 ----- 72
Query 121 APDNKAPGSTAPPAHGVTSADPTSPPGSTAPPAHGVTSADPTSPPGSTAPPAHGVTS 180
Db 73 ----- 72
```

Thu May 8 16:14:02 2003

us-09-658-621b-2.ram

Page 10

```
Qy 181 APDNRPALGSTAPVHVNVTSASGSASGASTLVHNGTSARATTPPASKSTPPSIPSHSD 240
Db 73 ----- 72
Qy 241 TPTTLASHSTKTDASSHSTHSTVPLTSSNHSTSPQLSTGVSPFPLSFHISNLQFNSLED 300
Db 73 -----LSTGVSPFPLSFHISNLQFNSLED 97
Qy 301 PSTDYVOELQRDISEMFLQIYKOGFLGLSNIKFRPGSVVWOLTLAFREGTINVHDETQ 360
Db 98 PSTDYVOELQRDISEMFLQIYKOGFLGLSNIKFRPGSVVWOLTLAFREGTINVHDETQ 157
Qy 361 FNOYKTEAASRYNLTSIDVSASHVPPFSAQSGAGVPGWGINLVLCVLAIVYLA 420
Db 158 FNOYKTEAASRYNLTSIDVSASHVPPFSAQSGAGVPGWGINLVLCVLAIVYLA 217
Qy 421 LAVQCCRRKNGQLDIPPARDTYHPMSXYPTYTHGRYPSPSTDSPFEKVSAGNGSS 480
Db 218 LAVQCCRRKNGQLDIPPARDTYHPMSXYPTYTHGRYPSPSTDSPFEKVSAGNGSS 277
Qy 481 LSYTNPAVAATSANL 495
Db 278 LSYTNPAVAATSANL 292
```

Search completed: May 1, 2003, 23:19:09
Job time : 281.036 secs

Thu May 8 16:14:01 2003

US-09-658-621b-2.rapb

Page 1

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:12:56 ; Search time 31.4286 Seconds
(without alignments)
1359.015 Million cell updates/sec

Title: US-09-658-621b-2

Perfect score: 2562

Sequence: 1 MPTGTGSPFFLLTLTVTV.....NGSSSLSTYNPAVAATSNL 495

Scoring table: BLOSUM62

Searched: 328255 seqs, 86286695 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubppa/PCT08_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2282.5	89.1	508	9 US-10-057-136-20	Sequence 20, App1
2	2130	83.1	1255	9 US-09-996-069-10	Sequence 10, App1
3	1640	64.0	321	10 US-09-925-301-861	Sequence 661, App
4	465	18.1	100	9 US-09-965-131-6	Sequence 6, App1
5	240	9.4	377	9 US-10-149-819-7	Sequence 7, App1
6	234	9.1	1162	10 US-09-745-008-34	Sequence 34, App1
7	233.5	9.0	1367	10 US-09-801-368-108	Sequence 108, App
8	230	9.0	554	9 US-09-996-069-3	Sequence 3, App1
9	230	9.0	957	10 US-10-025-280-1065	Sequence 1065, App
10	230	9.0	957	10 US-09-922-217-1065	Sequence 1065, App
11	230	9.0	957	10 US-09-833-263-1065	Sequence 1068, App
12	216.5	8.5	5179	10 US-10-025-380-1068	Sequence 1068, App
13	216.5	8.5	5179	10 US-09-922-217-1068	Sequence 1068, App
14	216.5	8.5	5179	10 US-09-833-263-1068	Sequence 1068, App
15	215	8.4	45	9 US-09-996-069-7	Sequence 7, App1
16	215	8.4	51	9 US-09-996-069-2	Sequence 2, App1
17	206.5	8.1	605	10 US-09-801-368-428	Sequence 428, App
18	205	8.0	40	9 US-09-996-069-9	Sequence 9, App1
19	205	8.0	46	9 US-09-996-069-5	Sequence 5, App1

20	205	8.0	1060	10 US-09-955-909-2	Sequence 2, App1
21	192	7.5	941	12 US-10-124-557-14	Sequence 14, App1
22	192	7.5	1022	12 US-10-124-557-84	Sequence 84, App1
23	192	7.5	1038	12 US-10-124-557-74	Sequence 74, App1
24	192	7.5	1049	12 US-10-124-557-58	Sequence 58, App1
25	192	7.5	1140	12 US-10-124-557-104	Sequence 104, App
26	192	7.5	1270	12 US-10-124-557-42	Sequence 42, App1
27	192	7.5	1311	12 US-10-124-557-44	Sequence 44, App1
28	192	7.5	1313	12 US-10-124-557-142	Sequence 142, App
29	192	7.5	1314	12 US-10-124-557-50	Sequence 50, App1
30	192	7.5	1320	12 US-10-124-557-46	Sequence 46, App1
31	192	7.5	1320	12 US-10-124-557-60	Sequence 40, App1
32	192	7.5	1354	12 US-10-124-557-48	Sequence 48, App1
33	192	7.5	1361	12 US-10-124-557-40	Sequence 40, App1
34	192	7.5	1363	12 US-10-124-557-52	Sequence 52, App1
35	192	7.5	1404	12 US-10-124-557-2	Sequence 2, App1
36	192	7.5	1404	12 US-10-124-557-62	Sequence 62, App1
37	191	7.5	386	10 US-09-864-761-35720	Sequence 35720, A
38	188.5	7.4	771	9 US-10-007-370-28	Sequence 28, App1
39	188.5	7.4	797	9 US-10-007-370-2	Sequence 2, App1
40	187.5	7.3	719	9 US-10-007-370-4	Sequence 4, App1
41	181.5	7.1	5877	9 US-10-142-515-11	Sequence 11, App1
42	180	7.0	699	9 US-10-121-988-143	Sequence 143, App
43	179	7.0	596	9 US-10-063-547-100	Sequence 100, App
44	179	7.0	596	9 US-10-174-590-310	Sequence 310, App
45	179	7.0	596	9 US-10-176-758-310	Sequence 310, App

ALIGNMENTS

RESULT 1
US-10-057-136-20
; Sequence 20, Application US/10057136
; Publication No. US20030021770A1
; GENERAL INFORMATION:
; APPLICANT: SCHLON, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUPFE, DONALD
; APPLICANT: PANICALI, DENNIS
; APPLICANT: GRITZ, LINDA
; TITLE OF INVENTION: TUMOR-ASSOCIATED ANTIGEN
; FILE REFERENCE: 700953/47113C
; CURRENT FILING DATE: 2002-01-25
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 09/366,670
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/038,253
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 20
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-136-20
Query Match 89.1% Score 2282.5 DB 9; Length 508;
Best Local Similarity 89.2% Pred. No. 5.4e-105;
Matches 455; Conservative 6; Mismatches 32; Indels 17; Gaps 3;
QY 1 MPTGTGSPFFLLTLTVTV.....TVVSGHASTPGEKETSATRRSSVPSSTK 51
DB 1 MPTGTGSPFFLLTLTVTVTVPAKAVVVGSGHASTPGEKETSATRRSSVPSSTK 60
QY 52 NAVSMTSVLSHSGSGSGSTGQCVTLAPATEPPASGAATWGDYVSVVPRPLSGT 111
DB 61 NAVSMTS--LISNGSVGSGTAPAPAGVTSABDTPAPAGSTAPAPAGVTSAPDRPAPGST 118
QY 112 TPAHVHTSAPDKKAPGATAPAPAGVTSABDTR-----PPGSTAPAAHGVTSAPDR 165

D b	119	APPAAGVTSADTRPPAGSSTAEPAGVTSAPDTRRAGDSTAPGSTAPPAHGVTSAPDTR	178
Q y	166	PAPGSTAPPAHGVTSAPDNRPALGSTAPVHNVTASGSASGSASTLVINGTSAAATTP	225
D b	179	PAPGSTAPPAHGVTSAPDNRPALGSTAPVHNVTASGSASGSASTLVINGTSAAATTP	238
Q y	226	AASKSTPPSPHSDPTTLAHSHTKTDASTHGVTPLTSSMHSSTPOLSTGVSEFFL	285
D b	239	AASKSTPPSPHSDPTTLAHSHTKTDASTHGVTPLTSSMHSSTPOLSTGVSEFFL	298
Q y	286	SPHINLTQFNSLEDPSTVYVOELORDISAMFLOIYKQGGFLGSHKRRPSSVVOGLT	345
D b	299	SPHINLTQFNSLEDPSTVYVOELORDISOMFLOIYKQGGFLGSHKRRPSSVVOGLT	358
Q y	346	APREGTINVHDETQNOQKTEAASRYNLTIIDVSVSHPPFPAQSAGVGWGMIALIV	405
D b	359	APREGTINVHDETQNOQKTEAASRYNLTIIDVSVSHPPFPAQSAGVGWGMIALIV	418
Q y	406	LVCVTLVALAIYTLAALAVCCCKRNKYQGLDIPARDYHPMSEPTVTHGRGVPSSTD	465
D b	419	LVCVTLVALAIYTLAALAVCCCKRNKYQGLDIPARDYHPMSEPTVTHGRGVPSSTD	478
Q y	466	RSPTKEVSAAGGSSLSYTNPAVAATSANI	495
D b	479	RSPTKEVSAAGGSSLSYTNPAVAATSANI	508

```

RESULT 2
US-09-996-069-10
: Sequence 10, Application US/09996069
: Publication No. US20030036199A1
: GENERAL INFORMATION:
: APPLICANT: Bamdad, Cynthia
: APPLICANT: Bamdad, R. Shoshana
: TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIB
: TITLE OF INVENTION: AND COMPOSITIONS AND METHODS FOR TREATMENT OF CANCER
: FILING REFERENCE: 001015/70071
: CURRENT APPLICATION NUMBER: US/09/996, 069
: CURRENT FILING DATE: 2001-11-27
: NUMBER OF SEQ ID NOS: 35
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 10
: LENGTH: 1255
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-996-069-10

```

	Query Match	83.1%;	Score 2130;	DB 9;	Length 1255;
	Best Local Similarity	39.0%;	Pred. No. 3,68-97;		
	Matches 489;	Conservative	1;	Mismatches 5;	Indels 760; Gaps 1;
Qy	1	MPRGQSPFFLLILLYLVMTVGTSGHNSSTPGGEKETAATQRSSVSGSTERNKVASMTSY	60		
Db	1	MPRGQSPFFLLILLYLVMTVGTSGHNSSTPGGEKETAATQRSSVSGSTERNKVASMTSY	60		
Qy	61	LSHSPGSGSSTTGQDQDYLAPATEPAGSGAATWGDVSVPTRRPALGSTAPPAAHGVTS	120		
Db	61	LSHSPGSGSSTTGQDQDYLAPATEPAGSGAATWGDVSVPTRRPALGSTAPPAAHGVTS	120		
Qy	121	APDNKPAGSTAPPAGVTSAPDTRPPGSGTAPAAHGVTSAPDCT-----	164		
Db	121	APDNKPAGSTAPPAGVTSAPDTRPPGSGTAPAAHGVTSAPDTRPPAGSTAPPAAHGVTS	180		
Qy	165	-----	164		
Db	181	APDTRPAPGSTAPPAGVTSAPDTRPPAGSTAPPAAHGVTSAPDTRPPAGSTAPPAAHGVTS	240		
Qy	165	-----	164		
Db	241	APDTRPAPGSTAPPAGVTSAPDTRPPAGSTAPPAAHGVTSAPDTRPPAGSTAPPAAHGVTS	300		
Qy	165	-----	164		

Dh	301	ADPTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	360
Qy	165	-----	164
Dh	361	ADPTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	420
Qy	165	-----	164
Dh	421	ADPTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	480
Qy	165	-----	164
Dh	481	ADPTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	540
Qy	165	-----	164
Dh	541	ADPTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	600
Qy	165	-----	164
Dh	601	ADPTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	660
Qy	165	-----	164
Dh	661	ADPTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	720
Qy	165	-----	164
Dh	721	ADPTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	780
Qy	165	-----	164
Dh	781	ADPTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	840
Qy	165	-----	164
Dh	841	ADPTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	900
Qy	165	-----	200
Dh	901	ADPTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	960
Qy	201	ASGASGASASTLWNHGTSAKATTTTPAKSKTPESIPSHHSDDPTTLASHSTYKTDASSTHS	260
Dh	961	ASGASGASASTLWNHGTSAKATTTTPAKSKTPESIPSHHSDDPTTLASHSTYKTDASSTHS	1020
Qy	261	TVPPITSSNNSTSPOLSTGVSFFPLSPHISNLOFNSSLEDPSTDYQELORDISEMFLOI	320
Dh	1021	SVPPITSSNNSTSPOLSTGVSFFPLSPHISNLOFNSSLEDPSTDYQELORDISEMFLOI	1080
Qy	321	YKQGGFTGLSNIKFRPGSVVQVLTLPFRGTTNVHDEVTOFNQYKTEAASNYNLTISDVS	380
Dh	1081	YKQGGFTGLSNIKFRPGSVVQVLTLPFRGTTNVHDEVTOFNQYKTEAASNYNLTISDVS	1140
Qy	381	VENHVPFPPSAQSAGVPEMGITALLIVTCLVALAIVTILALAVCQCRKRNKGQIDIFPAR	440
Dh	1141	VSDVFPFPPSAQSAGVPEMGITALLIVTCLVALAIVTILALAVCQCRKRNKGQIDIFPAR	1200
Qy	441	DTYHPMSEYTYTTHGKIVYPPSSITDRSPLEKVSAGNGSSLSLTYNDAVAASANL	495
Dh	1201	DTYHPMSEYTYTTHGKIVYPPSSITDRSPLEKVSAGNGSSLSLTYNDAVAASANL	1255
RESULT 3			
US-09-925-301-861			
Sequence 861, Application US/09925301			
Patent No. US20020052308A1			
GENERAL INFORMATION:			
APPLICANT: Kosen et al.			
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies			
FILE REFERENCE: P106			
CURRENT APPLICATION NUMBER: US/09/925_301			

CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 861
LENGTH: 321
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: 1181
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-861

Query Match 64.0%; Score 1640; DB 10; Length 321;
Best Local Similarity 99.1%; Pred. No. 1e-73;
Matches 318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 175 AAGVTAPDNRPALGSTAPVHNVTSASGASGASTLVHNGTSAPATTPACKSTPPI 234
DB 1 AAGVTAPDNRPALGSTAPVHNVTSASGASGASTLVHNGTSAPATTPACKSTPPI 60
QY 235 PSHHSDPTTASHSTKTDASSTHSTVPPPLTSSNHSTSPOLSTGVSEFFLSFHSINLOF 294
DB 61 PSHHSDPTTASHSTKTDASSTHSTVPPPLTSSNHSTSPOLSTGVSEFFLSFHSINLOF 120
QY 295 NSSLEDPSTDYOELODISEMFLQIKYKGGFLGLSNIKRPGSVVVOULTAPREGTINV 354
DB 121 NSSLEDPSTDYOELODISEMFLQIKYKGGFLGLSNIKRPGSVVVOULTAPREGTINV 180
QY 355 HDVETQNGYKTEAASRYNTTIDSVSHVPEPPSPAGSGAGVGGIALLVLVCVVALA 414
DB 181 HDVETQNGYKTEAASRYNTTIDSVSHVPEPPSPAGSGAGVGGIALLVLVCVVALA 240
QY 415 IVVLILAACCCRRKRGOLDIPADDTYHMSRYPTTHHGRGVPPSTGRSPYEVKVA 474
DB 241 IVVLILAACCCRRKRGOLDIPADDTYHMSRYPTTHHGRGVPPSTGRSPYEVKVA 300
QY 475 GNGGSLSTTPNPAVATSANL 495
DB 301 GNGGSLSTTPNPAVATSANL 321

RESULT 4
US-09-965-131-6
Sequence 6; Application US/09965131
Patent No. US20020160502A1
GENERAL INFORMATION:
APPLICANT: Chung, Maureen A.
APPLICANT: Sharma, Surendra
APPLICANT: Chang, Helena R.
APPLICANT: O'Donnell, Mark A.
TITLE OF INVENTION: RECOMBINANT BCG VACCINES FOR THE
FILE REFERENCE: WI-014CP
CURRENT APPLICATION NUMBER: US/09/965,131
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/235,455
PRIOR FILING DATE: 2000-09-26
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PaasSeq for Windows Version 4.0
SEQ ID NO: 6
LENGTH: 100
TYPE: PRT
ORGANISM: Homo sapiens
US-09-965-131-6

Query Match 18.1%; Score 465; DB 9; Length 100;
Best Local Similarity 87.9%; Pred. No. 1.2e-16;
Matches 87; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 98 VTSVPVTRPALGSTTPPAHDVTSAPDNKAPGSTAPPAHGVTSAPDTRPPPGSTAPAAHG 157
DB 2 VTSAPDTRPPAPGSTAPPAHGVTSAPDTRPPAPGSTAPPAHGVTSAPDTRPPAPGAHG 61
QY 158 VTSAPDTRPPAPGSTAPPAHGVTSAPDNKAPGSTAPPAH 196
DB 62 VTSAPDTRPPAPGSTAPPAHGVTSAPDTRPPAPGSTAPPAH 100

RESULT 5
US-10-149-819-7
Sequence 7; Application US/10149819
Publication No. US20030044913A1
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: YUE Henry
APPLICANT: AZIMZAI Yalda
APPLICANT: TANG Y. Tom
APPLICANT: PATTERSON, Chandra
APPLICANT: BAIGER, Maria R.
APPLICANT: LU, Dying Aina M.
APPLICANT: SHAH, Preeti
APPLICANT: LAU, Preeti
APPLICANT: AU-YOUNG, Janice
APPLICANT: BURFORD, Neil
TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
FILE REFERENCE: PF-0760 PCT
CURRENT APPLICATION NUMBER: US/10/149,819
CURRENT FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
PRIOR FILING DATE: 1999-12-10; 1999-12-16
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PERL Program
SEQ ID NO: 7
LENGTH: 377
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030044913A1 1825182CD1
US-10-149-819-7

Query Match 9.4%; Score 240; DB 9; Length 377;
Best Local Similarity 30.5%; Pred. No. 4.4e-05;
Matches 90; Conservative 44; Mismatches 91; Indels 70; Gaps 17;

QY 37 TGAIGRSVSGSTKNAVMTSSVUSHSRSGSSSTGCGDVTIAPATPAGSASATMGQ 96
DB 118 SASRTITTLFNVTF--FLPQNAFTTSSRENVNTSSV--ATLAPVNSAP-----Q 163
QY 97 DVTSVPTTRPALGSTTPPAHDVTSAPDNKAPGST--APPAHGVTSAPDTRPPPGSTAPA 154
DB 164 DTTAAP--PTPSATTP-----APPSSAPPEITTAAP--TPSATTOAPSSSAP 209
QY 155 AAGVTAPDTRPA-----GSTAPPAHGVTSAPDNKPA-----LGSTAPVHNVTSASGS 204
DB 210 E--TTAAPPTTPATTPAPSSSAPPE--TTAAPPTPSATTPAPSSSAP-----PTTASH 248
QY 205 AAGSAGTIVHNGTSAPATTT-----PASKSTPPSIPSHSDT-----PTTASH 248
DB 262 PTPSATTTDPSSASAPPEITTAAPPTPSATTPAPSSSAPPAOETTAAPPTTNSPTTAAVD 321
QY 249 STKTDASSTHSTVPPPLTSSNHSTSPOLSTG--VSFFPLSFHSINLOFNSILED 300
DB 322 TSETSAAPHTGTTTSVTTGTTTQPTSPAGQNKISRFL--YMKNI--LRIID 373

RESULT 6
US-09-745-008-34
Sequence 34; Application US/09745008
Patent No. US20020137667A1
GENERAL INFORMATION:

```

; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Mircelo A.
; TITLE OF INVENTION: T. Cruz1-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,981
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
US-09-745-008-34

```

```

Query Match          9.1%; Score 234; DB 10; Length 1162;
Best Local Similarity 28.7%; Pred. No. 0.00026;
Matches 87; Conservative 34; Mismatches 134; Indels 48; Gaps 11;

```

```

QY 16 TVLFTVYTGSGH-----ASTPGKEKESATQSSVPSSTENKAYSMTSVLSHS 65
DB 707 TPSTPVDSSAHGTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHG 766
QY 66 PGSSSTTQGDVTLAPATEPAGSAT-----KGQDVTSVPTRPALGSTTP---AHD 117
DB 767 TPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHG 826
QY 118 VTSA-PDKRPAGSTAP---AHGVTAP-DTRPPPGSTAP---AHGVTAPDTRPAPG 169
DB 827 TPSTPVDSSAHGTPSTPADSSAHSTPSTPADSSAHGTPSTPADSSAHG 886
QY 170 STAPP-----AHGVTAPDKRPALGSTAPPVHNTVTSAGSAGSASTLV 213
DB 887 TPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHGTPSTVD--SSAKTSTPADSSA 944
QY 214 HNGTSAPATTPPAKSTPPSPSHS--DTPPTLASHST-KTDASTHSTHST-VPPLTSSN 269
DB 945 HSTSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSA 1004
QY 270 HST 272
DB 1005 HST 1007

```

```

RESULT 7
US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Paul
; APPLICANT: Silva, Jeli
; APPLICANT: Summerville, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440

```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 1367
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

```

```

Query Match          9.1%; Score 233.5; DB 10; Length 1367;
Best Local Similarity 28.9%; Pred. No. 0.00032;
Matches 88; Conservative 45; Mismatches 123; Indels 49; Gaps 12;

```

```

QY 20 VYTGSGHASTPGKEKESATQSSVPSSTENKAYSMTSVLSHS---PGSSSTT 73
DB 344 VTSSTESSAPVPTPSSSTESSAPVTSSTESSAPVTSSTESSAPVPTPSS 403
QY 74 QGDVTLAPATEPAGSATWG-ODVTSVPTRPALGSTTPPAHDVTSAPDKPAP--- 128
DB 404 ESSAPVTSSTESSAPVTSSTESSAPVTSSTESSAPVTSSTESSAPVPTPSS 463
QY 129 ---GSTAPPAHGYT---SAPDTRPPGSTAPPAHGYTAPDTR---PAGSTAPPAH 176
DB 464 STTESSAPVTSSTESSAPVPTPSSSTESSAPVTSSTESSAPVTSSTESSAPVPTPSS 522
QY 177 GVTSA-PDKRPALGST---APVHNTVTSAGSA---SGASITVHNGTSAPATTPPAKS- 229
DB 523 ---SSAPATVTSSTESSAPVTSSTESSAPVPTPSSSTESSAPVTSSTESSAPVPTPSS 580
QY 230 --TPSPISHSHTPTTLASHST-----KTDASTHSTHSTVPPLTSSNHST--- 273
DB 581 VPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPVTSSTESSAPVTSSTESSAPVPT 640
QY 274 POLST 278
DB 641 PSSST 645

```

```

RESULT 8
US-09-996-069-3
; Sequence 3, Application US/09996069
; Publication No. US20030036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia
; APPLICANT: Bamdad, R. Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIB
; TITLE OF INVENTION: AND COMPOSITIONS AND METHODS FOR TREATMENT OF CANCER
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996,069
; PRIOR FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-3

```

```

Query Match          9.0%; Score 230; DB 9; Length 54;
Best Local Similarity 95.8%; Pred. No. 2.1e-05;
Matches 46; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 341 VQTLAFREGTINVDVETQFNQYKTEASRYNLTISDVSYSVHPFPF 388
DB 1 VQTLAFREGTINVDVETQFNQYKTEASRYNLTISDVSYSVHPFPF 48

```

```

RESULT 9
US-10-025-380-1065
; Sequence 1065, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshan
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secret, Heather

```

APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tonglong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick Thomas S.
APPLICANT: Carter, Darick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025.380
CURRENT FILING DATE: 2001-12-19
NUMBER OF SEQ ID NOS: 1123
SOFTWARE: fastseq for Windows Version 4.0
SEQ ID NO 1065
LENGTH: 957
TYPE: PRT
ORGANISM: Homo sapiens
US-10-025-380-1065

Query Match 9.0%; Score 230; DB 9; Length 957;
Best Local Similarity 22.3%; Pred. No. 0.00033;
Matches 131; Conservative 74; Mismatches 210; Indels 172; Gaps 23;

28 SSTPGKEKESATQSSVPSSTKNAVMTSSVL-----SSH-SPGS-----GSTTQ 75
205 SUTTSRGESTTSHSST-THTISAPETTSALVEPTSYHSSPGSTATTNPFDSSTSG 263
QY 76 --QDVTLAAPTEPASGAATWGQDVTY-----PYTRPAL-----GSTTP-AND 117
DB 264 RSEESTASHNQATITVLPARSTSVLGEESTSPISSGMETALPSTITPGLSRK 323
QY 118 VTSAPDNKPAAGSTAPPAAGVTSADTRPPPGSTAPAAAGVTSADTRPPAG----- 169
DB 324 STPHSRPSPATTLSPASTTSSGVSEESTSHSRPGSTHTTAFFDSTTTPGLSRKSTTS 383
QY 170 -----STAPPAGVTSAPDNRPALGSTAPVHNVTASGS----- 204
DB 384 HSSPGSTDTTLPLASTTSSGVEESTTSHSPGSTDTLSPGSTTALSFGQESTTFHSP 443
QY 205 -----ASG--SASTLVHNGT-SARATTPPAKSTP-----PSIPSHNDTP 242
DB 444 GSTHTTLFPDSTTSSGIVEASTRVHSSGTSPRTTLPASTSPGLQESTAPFOTHPASTH 503
QY 243 TFLASHSTKT---DASTHSTVPLTSSNHSSTSPOLSTGVFFFLPHISNLOFNSSLE 299
DB 504 TTPSPATAPAVEESTTYHRS-PSSTPTTHFPASSTSGHSEKSTTFHSP---DASGT 559
QY 300 DSGTYYOELORDISEMFIQIKQGFGLSNIKFRPGSVVVCULTAFREGTINVDVET 359
DB 560 TSSAHSTSGRES-----TTSKRISPGSTEL-TTLGSTTTGGLSEAST 603
QY 360 QF-----NOYKTEASRYNLTISDVSVSHVPPPSAOSGAGV-- 396
DB 604 TYSSPSPPTTLSPASMTSLGVGEESTSRQGSTHVSIPASTTTPGLSEESTTVTS 663
QY 397 --PGMGIALVLVCVLAVALAIVYLLALAVCCQRRKNYQGLDIFPARDTYHMESEYF-TYH 453
DB 664 SSPG-----STETTVFPSTTTSVKGEEPTTFH 691
QY 454 -----THGRVPPSTDRSPYEKVSAGNGSSLSYTNPAVAATSNL 495
DB 692 SRPASTHTTLFTEDSTTSGLTREASTAPGSPASTGTGLPATLTADL 738

RESULT 10
US-09-922-217-1065

Sequence 1065, Application US/09922217
Patent No. US2002007641A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tonglong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole Lynn
APPLICANT: King, Gordon E.
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasir A. W.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C13
CURRENT APPLICATION NUMBER: US/09/922.217
CURRENT FILING DATE: 2001-08-03
NUMBER OF SEQ ID NOS: 1124
SOFTWARE: fastseq for Windows Version 4.0
SEQ ID NO 1065
LENGTH: 957
TYPE: PRT
ORGANISM: Homo sapiens
US-09-922-217-1065

Query Match 9.0%; Score 230; DB 10; Length 957;
Best Local Similarity 22.3%; Pred. No. 0.00033;
Matches 131; Conservative 74; Mismatches 210; Indels 172; Gaps 23;

28 SSTPGKEKESATQSSVPSSTKNAVMTSSVL-----SSH-SPGS-----GSTTQ 75
205 SUTTSRGESTTSHSST-THTISAPETTSALVEPTSYHSSPGSTATTNPFDSSTSG 263
QY 76 --QDVTLAAPTEPASGAATWGQDVTY-----PYTRPAL-----GSTTP-AND 117
DB 264 RSEESTASHNQATITVLPARSTSVLGEESTSPISSGMETALPSTITPGLSRK 323
QY 118 VTSAPDNKPAAGSTAPPAAGVTSADTRPPPGSTAPAAAGVTSADTRPPAG----- 169
DB 324 STPHSRPSPATTLSPASTTSSGVSEESTSHSRPGSTHTTAFFDSTTTPGLSRKSTTS 383
QY 170 -----STAPPAGVTSAPDNRPALGSTAPVHNVTASGS----- 204
DB 384 HSSPGSTDTTLPLASTTSSGVEESTTSHSPGSTDTLSPGSTTALSFGQESTTFHSP 443
QY 205 -----ASG--SASTLVHNGT-SARATTPPAKSTP-----PSIPSHNDTP 242
DB 444 GSTHTTLFPDSTTSSGIVEASTRVHSSGTSPRTTLPASTSPGLQESTAPFOTHPASTH 503
QY 243 TFLASHSTKT---DASTHSTVPLTSSNHSSTSPOLSTGVFFFLPHISNLOFNSSLE 299
DB 504 TTPSPATAPAVEESTTYHRS-PSSTPTTHFPASSTSGHSEKSTTFHSP---DASGT 559
QY 300 DSGTYYOELORDISEMFIQIKQGFGLSNIKFRPGSVVVCULTAFREGTINVDVET 359
DB 560 TSSAHSTSGRES-----TTSKRISPGSTEL-TTLGSTTTGGLSEAST 603
QY 360 QF-----NOYKTEASRYNLTISDVSVSHVPPPSAOSGAGV-- 396
DB 604 TYSSPSPPTTLSPASMTSLGVGEESTSRQGSTHVSIPASTTTPGLSEESTTVTS 663
QY 397 --PGMGIALVLVCVLAVALAIVYLLALAVCCQRRKNYQGLDIFPARDTYHMESEYF-TYH 453
DB 664 SSPG-----STETTVFPSTTTSVKGEEPTTFH 691
QY 454 -----THGRVPPSTDRSPYEKVSAGNGSSLSYTNPAVAATSNL 495
DB 692 SRPASTHTTLFTEDSTTSGLTREASTAPGSPASTGTGLPATLTADL 738

```
RESULT 11
US-09-833-263-1065
Sequence 1065, Application US/09833263
GENERAL INFORMATION:
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Stolck, John A.
APPLICANT: Meagher, Madeleine J.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
FILE REFERENCE: 210121.471C12
CURRENT APPLICATION NUMBER: US/09/833.263
NUMBER OF SEQ ID NOS: 1093
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1065
LENGTH: 957
TYPE: PRF
ORGANISM: Homo sapiens
US-09-833-263-1065

Query Match          9.0%; Score 230; DB 10; Length 957;
Best Local Similarity 22.3%; Pred. No. 0.00033;
Matches 131; Conservative 74; Mismatches 210; Indels 172; Gaps 23;

QY 28 SSTPGEKSTATQSSVPSSTKNAVMTSSVL-----SSH-SRGS-----GSTTQ 75
DB 205 SQTSGRGESTSHSSST-THTTSSAPSTTSLMVEPSTHSGSTATHTPPDSSTSG 263
QY 76 --QDVTAPATPAPASGAATWGDVTSV-----PYTRPAL-----GSTTP--AHD 117
DB 264 RSEESTSHNDATGTVLPAPASTTSLVLSGSTTSPSSGSEMTALPAGSTTTPGLSEK 323
QY 118 VTSAADNKPAPASTAPPAHGVTSAPDTPPPGSTAPAAHGVTSAPTPRPAPG----- 169
DB 324 STTHSSPSPSPATITLSPASTTSGVSEESTTSHSPGSGTMTAFDSTTTTGLSRHSTTS 383
QY 170 -----STAPPAHGVTSAPDNKPALGSTAPPVNVTSAAGS----- 204
DB 384 HSPSGSTDTLLPASTTSGPQSGESTTSHSSPGSTDTLSPGSTALSPQSGESTTFFHSSP 443
QY 205 --S-----ASG--SASTLVHNGT--SARATTPPAKSTP-----PSIPSHHGT 242
DB 444 GSTTTLFPDSTTSGGIVEASTRVHSGTSPPTLSPASTSPQLQGEFSTARQHPASTH 503
QY 243 TTLSHSTYKT--DASTHSTVPLTSSNHSTSPQSLSTGVSEFFLSFPHISMLQPNSTLE 289
DB 504 TTPTSTPTATAPVEESTTYHNS--PSSPTTPHPASTTSGHSEKSTTFHSSP--DASGT 559
QY 300 DPSDYYQGLQDISEMFLQYKGGFLGSLNIKERPGSVVYQULTAPREGTINHYET 359
DB 560 TFSASHSTTSGKGS-----TSKLSFGSTET--TLPGSTTTPGLSENST 603
QY 360 QF-----NQTETAAGRVYMLTSDVSVSHVPPFPAQSGAGY-- 396
DB 604 TFYSSPSPSTTTLSPASTSLGVEESTTYSRQPSGTHSTVSPASTTTPGLSEESTTVYS 663
QY 397 --PGWGIALLVVAVVLAIVLIALAVQCQRKQVQGLDIPPADTYHPSSEY--TYH 453
DB 664 SSPG--STETVFPFRSTTSSVGESEPTTFH 691
QY 454 -----THGRVPSSTDRSPYKVSAGNGGSSLSYTNPAVATSAHL 495
DB 692 SRPASTHTTTLFTEDSTYGLTEBSTAPPGSPASTOTGAPATLTADL 738

RESULT 12
US-10-025-380-1068
Sequence 1068, Application US/10025380
GENERAL INFORMATION:
APPLICANT: Xu, Jianshun
```

```
APPLICANT: Lodes, Michael J.
APPLICANT: Secret, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolck, John A.
APPLICANT: Wang, Tonglong
APPLICANT: Jiang, Yudi
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yael A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick Thomas S.
APPLICANT: Carter, Darick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025,380
NUMBER OF SEQ ID NOS: 1129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1068
LENGTH: 5179
TYPE: PRF
ORGANISM: Homo sapiens
US-10-025-380-1068

Query Match          8.5%; Score 216.5; DB 9; Length 5179;
Best Local Similarity 26.0%; Pred. No. 0.0079;
Matches 82; Conservative 46; Mismatches 126; Indels 61; Gaps 13;

QY 2 TPTQSPFLLILLITLVVYTGSGHASTPGEKETSATQNSV-----PSTKNAVMT 57
DB 1478 SPPTTP-----SPPTTTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTP 1530
QY 58 SSVLSSHSPGSGS--TQSGDVTLPAPATPAPASGAATWGDVTSVPVTRPALGSTT-- 112
DB 1531 TTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTP 1586
QY 113 -PPADVTAP--DNKPAPGSTAPPAHGVTSAPDTPR--PPGSTA-----PAA 155
DB 1587 SPPTTTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTP 1646
QY 156 HGVTSAPDTPAPGST--APPAHGVTSAP--DNRPALGSTAPPVNVTSAAGSAGS 208
DB 1647 TTTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTP 1706
QY 209 ASPLVHNGTSARATTPPAKSTPSPISPHSHDPTTLASHSTKTDAS--STHSTVPLT 266
DB 1707 -----SPITTTTPSSSTTP-----SPPTTMTTPSPPTTPSPPTTPMTLPTT 1751
QY 267 SSNHSTSPQSLSTGVS 281
DB 1752 TSSPTTTPDPPSIT 1766

RESULT 13
US-09-922-217-1068
Sequence 1068, Application US/09922217
GENERAL INFORMATION:
APPLICANT: Xu, Jianshun
APPLICANT: Lodes, Michael J.
APPLICANT: Secret, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolck, John A.
APPLICANT: Wang, Tonglong
APPLICANT: Jiang, Yudi
APPLICANT: Smith, Carole Lynn
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
```



```

; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-922-217-1068

Query Match
Best Local Similarity 26.0%; Pred. No. 0.0079;
Matches 82; Conservative 46; Mismatches 126; Indels 61; Gaps 13;

QY 2 TPGTGSPPFLLLTLVTVTSGHASSTPGSEKETSATQSSV---PSSTKKNVSM 57
DB 1478 SPTTTT-----SPTTTTTPPTTTPSPMTTITTPASTTTLPTTTPSPPTTT 1530
QY 58 SSVLSHSPGSGSS---TTGGQDVLAPATBPASGSAATWGDVTSVPTRPALGTT-- 112
DB 1531 TTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1586
QY 113 -PPADVTSAP---DNKPAAGSTAPPAHGVTSAPDTRP--PGSTA-----PAA 155
DB 1587 SPTTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1646
QY 156 HGVTSAPDTRPAPGST---APPAHGVTSAP---DNKPAAGSTAPPAHNVTSASGSGS 208
DB 1647 TTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1706
QY 209 ASTLVNNGSAPATTPPASKTSPSPSHSDPTPTLASHTKTDAS--STHASTVPLT 266
DB 1707 -----SPTTTTTPSSTTP-----SPTTMTTPSPPTTTPPTTTPPTT 1751
QY 267 SSNHSSTPQLSTGVS 281
DB 1752 TSSPLTTPPLPSPIT 1766

RESULT 14
US-09-833-263-1068
; Sequence 1068, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-833-263-1068

Query Match
Best Local Similarity 8.5%; Score 216.5; DB 10; Length 5179;
Matches 82; Conservative 46; Mismatches 126; Indels 61; Gaps 13;

QY 2 TPGTGSPPFLLLTLVTVTSGHASSTPGSEKETSATQSSV---PSSTKKNVSM 57
DB 1478 SPTTTT-----SPTTTTTPPTTTPSPMTTITTPASTTTLPTTTPSPPTTT 1530
```

```

QY 58 SSVLSHSPGSGSS---TTGGQDVLAPATBPASGSAATWGDVTSVPTRPALGTT-- 112
DB 1531 TTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1586
QY 113 -PPADVTSAP---DNKPAAGSTAPPAHGVTSAPDTRP--PGSTA-----PAA 155
DB 1587 SPTTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1646
QY 156 HGVTSAPDTRPAPGST---APPAHGVTSAP---DNKPAAGSTAPPAHNVTSASGSGS 208
DB 1647 TTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1706
QY 209 ASTLVNNGSAPATTPPASKTSPSPSHSDPTPTLASHTKTDAS--STHASTVPLT 266
DB 1707 -----SPTTTTTPSSTTP-----SPTTMTTPSPPTTTPPTTTPPTT 1751
QY 267 SSNHSSTPQLSTGVS 281
DB 1752 TSSPLTTPPLPSPIT 1766

RESULT 15
US-09-996-069-7
; Sequence 7, Application US/09996069
; Patent No. US20030036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia
; APPLICANT: Bamdad, R. Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIB
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996.069
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 7
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-996-069-7

Query Match
Best Local Similarity 8.4%; Score 215; DB 9; Length 45;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 350 GTINVDVETOPNOYKTEAASRYNLTISDVSVSHVPPFSAQGA 394
DB 1 GTINVDVETOPNOYKTEAASRYNLTISDVSVSHVPPFSAQGA 45

Search completed: May 1, 2003, 23:21:43
Job time : 40.4286 secs
```


GenCore version 5.1.4.DS_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 22:59:10 ; Search time 63.8393 Seconds
(without alignments)
1033.205 Million cell updates/sec

Title: US-09-658-621b-2
Perfect score: 2562
Sequence: 1 MPTGQSPFFILLTLTVLTV.....NGSSLSYTPAVVATSNL 495

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2562	100.0	495	22 AAU00539	Human MUC1 polypep
2	2413	94.2	475	22 AAU00573	Human MUC1 polypep
3	2413	94.2	475	23 ABB77476	Human MUC1 SEQ ID
4	2391	93.3	473	22 AAE09508	Human mucin-1 (MUC
5	2312	90.2	455	21 AA171024	Human mucin-1 (MUC
6	2311.5	90.2	577	21 AA171030	Ubiquitin-E. coli
7	2295	89.6	455	19 AAAR33973	Transmembrane form
8	2282.5	89.1	508	19 AAAR7233	Mitochondrial protein
9	1777	69.4	348	13 AAAR7662	C-terminal region
10	1671	65.2	327	16 AAAR6298	Glycoprotein 39 C

11	1640	64.0	321	21 AAB3416	Human cancer assoc
12	1587	61.9	312	23 AAB4810	Human MUC1R consen
13	1587	61.9	312	23 AAB4810	Secreted form of H
14	1352	52.8	287	13 AAR7665	Mucin-derived prot
15	1273	49.7	273	17 AAR9418	Mucin-derived prot
16	1258.5	49.1	282	17 AAR9419	Mucin-derived prot
17	1174	45.8	255	17 AAR9420	Mucin-derived prot
18	1159.5	45.3	264	17 AAR9421	Mucin-derived prot
19	1091.5	42.6	240	17 AAR9422	Mucin-derived prot
20	1086.5	42.4	239	17 AAR9423	Mucin-derived prot
21	1082.5	42.3	338	21 AAV1028	Ubiquitin-E. coli
22	1076	42.0	216	21 AAV1022	Human Mucin 1 (MUC
23	863	33.7	295	21 AAV1027	Human Mucin 1 (MUC
24	854	33.3	173	21 AAV1021	Human Mucin 1 (MUC
25	834	32.6	180	13 AAR7664	C-terminal region
26	822	32.1	167	13 AAR7661	N-terminal region
27	822	32.1	167	13 AAR7663	N-terminal region
28	645	25.2	129	12 AAU83110	Human polymorphic
29	629	24.6	125	23 AAU84809	Human MUC1F consen
30	532.5	20.8	216	21 AAV92665	MUC-1 analogue con
31	488	19.0	215	21 AAU85008	Human melanocyte d
32	486	19.0	93	21 AAV1029	Ubiquitin-E. coli
33	482	18.8	105	16 AAR8022	Human Mucin 1 (MUC
34	482	18.8	105	16 AAR8022	Mucin peptide p105
35	465	18.1	100	23 AAB76181	Synthetic muc-1 pe
36	353	13.8	71	22 AAB76181	Glutathione-S-tran
37	353	13.2	177	20 AAB09548	Human mucin-1 (MUC
38	339	13.2	177	20 AAV29896	Human SDF1 and hum
39	337	13.2	69	21 AAV1020	Human Mucin 1 (MUC
40	336.5	13.1	191	21 AAV1026	Ubiquitin-E. coli
41	334	13.0	172	20 AAV29895	Human MUC and huma
42	283	11.0	60	16 AAR6297	Glycoprotein 39 N
43	266	10.4	95	17 AAR9426	Mucin-derived prot
44	262	10.2	76	17 AAR9424	Mucin-derived prot
45	251.5	9.8	105	17 AAR9427	Mucin-derived prot

ALIGNMENTS

RESULT 1					
ID	AAU00539	standard; Protein; 495 AA.			
AC	AAU00539;				
XX					
DT	12-SEP-2001 (first entry)				
XX					
DB	Human MUC1 polypeptide.				
XX					
XX	Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-I;				
KW	glycoprotein; cytotoxic T lymphocytes; T cell response; cancer; vaccine;				
KW	cancer gene therapy; diagnosis; treatment; inflammatory disorder;				
KW	organ transplant rejection; graft versus host disease.				
OS	Homo sapiens.				
XX					
XX	MO200118035-A2.				
PD	15-MAR-2001.				
XX					
PF	07-SEP-2000; 2000MO-EP08761.				
XX					
PR	08-SEP-1999; 99GB-0021242.				
PR	10-SEP-1999; 99EP-0402237.				
XX	03-MAR-2000; 2000US-0187215.				
XX					
PA	(TRGE) TRANSGENE SA.				
XX	(IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.				
XX	Taylor-Papadimitriou J, Heukamp IC, Offringa R, Melief CJM;				
PI	Actes B, Thomas M;				
XX					

DR WP1, 2001-235187/24.
 DR N-PSDB; AAS00565.
 XX New antigenic polypeptides of MUC-1 protein which activate cytotoxic T
 PT lymphocyte proteins and their analogues, useful for identifying a major
 PT histocompatibility complex class I restricted T cell response and for
 PT diagnosing cancer -
 XX
 XX
 PS Claim 1, Page 68-69; 81pp; English.
 CC The sequence represents a human MUC1 polypeptide. Derivative antigenic
 CC peptides of this protein bind at least one major histocompatibility
 CC complex class I (MHC-I) glycoprotein, which activates cytotoxic T
 CC lymphocytes to induce a protective response against tumours. Diagnosis of
 CC cancer involves determining the presence or absence in a host cell of MHC
 CC class I restricted T cell response to a MUC1 derivative, where the
 CC presence of the MHC class I restricted T cell response indicates that the
 CC host has cancer. Measurement of the level of MHC class I restricted T
 CC cell response is also useful to monitor the severity of cancer, a larger
 CC response indicating a more severe cancer. MUC1 derivatives are useful in
 CC cancer therapy and to follow MUC1 specific immune responses in patients
 CC during the course of disease and/or treatment. MUC1 DNA is useful in
 CC cancer gene therapy, vaccination and diagnosis. Compositions of the
 CC sequences are used in vaccines and treatments against cancer or diseases
 CC caused by an immune response, such as an inflammatory disorder, organ
 CC transplant rejection or graft versus host disease.
 XX
 SQ Sequence 495 AA:
 Query Match 100.0%; Score 2562; DB 22; Length 495;
 Best Local Similarity 100.0%; Pred. No. 5.8e-152;
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTPQTOSPPFLILLITVLTVTGSGHASTPGGKERTSATORSVPSSTENKNAVSMTSSV 60
 DB 1 MTPQTOSPPFLILLITVLTVTGSGHASTPGGKERTSATORSVPSSTENKNAVSMTSSV 60
 QY 61 LSHSPGSGSSTTQGDVTLAPATEPASGSAATWGDVTSVPTTRPALGSTTPPADVTS 120
 DB 61 LSHSPGSGSSTTQGDVTLAPATEPASGSAATWGDVTSVPTTRPALGSTTPPADVTS 120
 QY 121 APDKAPAPSTAPPAHGVTSAPTRPPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180
 DB 121 APDKAPAPSTAPPAHGVTSAPTRPPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180
 QY 181 APDKAPAPSTAPPAHGVTSAPTRPPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180
 DB 181 APDKAPAPSTAPPAHGVTSAPTRPPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180
 QY 241 TPTTLASHTKTDASTHSTHSTVPTLSSNHSSTQSLSTGVSPRISPRISMLOPNSLED 300
 DB 241 TPTTLASHTKTDASTHSTHSTVPTLSSNHSSTQSLSTGVSPRISPRISMLOPNSLED 300
 QY 301 PSTDIYQEIQRDISMFLQIKGGSTLGSINIKRPPGVVVOGLTAFREGSTINVHDEVTO 360
 DB 301 PSTDIYQEIQRDISMFLQIKGGSTLGSINIKRPPGVVVOGLTAFREGSTINVHDEVTO 360
 QY 361 PNOYKTEBASRYNLITDSVSVHPPPPASGAGVFGKIALVLCVLAIAIYVLLA 420
 DB 361 PNOYKTEBASRYNLITDSVSVHPPPPASGAGVFGKIALVLCVLAIAIYVLLA 420
 QY 421 LAVQCGRKNYQGLDIPAPDYYHPMSEVPTVTHGAYVPSSTDRSPYKRVASNGGSS 480
 DB 421 LAVQCGRKNYQGLDIPAPDYYHPMSEVPTVTHGAYVPSSTDRSPYKRVASNGGSS 480
 QY 481 LSYTNPAVAATSANT 495
 DB 481 LSYTNPAVAATSANT 495
 RESULT 2
 AAU00573 standard; Protein; 475 AA.

XX AC AAU00573;
 XX 12-SEP-2001 (first entry)
 DT
 XX
 XX Human MUC1 polypeptide variant.
 DE
 XX
 XX Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-I;
 KW glycoprotein; cytotoxic T lymphocytes; T cell response; cancer; vaccine;
 KW cancer gene therapy; diagnosis; treatment; inflammatory disorder;
 KW organ transplant rejection; graft versus host disease.
 XX
 XX Homo sapiens.
 OS
 XX WO200118035-A2.
 XX 15-MAR-2001.
 XX 07-SEP-2000; 2000WO-EP08761.
 PF 08-SEP-1999; 99GB-0021242.
 XX 10-SEP-1999; 99EP-0402237.
 PR 03-MAR-2000; 2000US-0187215.
 XX
 XX (TRGE) TRANSGENE SA.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
 XX
 XX Taylor-Papadimitriou J, Heukamp LC, Offringa R, Melief CJM;
 PI Acres B, Thomas M;
 PI
 XX
 DR WP1; 2001-235187/24.
 XX
 PT New antigenic polypeptides of MUC-1 protein which activate cytotoxic T
 PT lymphocyte proteins and their analogues, useful for identifying a major
 PT histocompatibility complex class I restricted T cell response and for
 PT diagnosing cancer -
 XX
 XX
 PS Disclosure; Fig 12; 81pp; English.
 XX
 XX The sequence represents a human MUC1 polypeptide variant. Derivative
 CC antigenic peptides of MUC1 protein bind at least one major
 CC histocompatibility complex class I (MHC-I) glycoprotein, which activates
 CC cytotoxic T lymphocytes to induce a protective response against tumours.
 CC Diagnosis of cancer involves determining the presence or absence in a
 CC host cell of MHC class I restricted T cell response to a MUC1 derivative,
 CC where the presence of the MHC class I restricted T cell response
 CC indicates that the host has cancer. Measurement of the level of MHC class
 CC I restricted T cell response is also useful to monitor the severity of
 CC cancer, a larger response indicating a more severe cancer. MUC1
 CC derivatives are useful in cancer therapy and to follow MUC1 specific
 CC immune responses in patients during the course of disease and/or
 CC treatment. MUC1 DNA is useful in cancer gene therapy, vaccination and
 CC diagnosis. Compositions of the sequences are used in vaccines and
 CC treatments against cancer or diseases caused by an immune response, such
 CC as an inflammatory disorder, organ transplant rejection or graft versus
 CC host disease.
 XX
 SQ Sequence 475 AA:
 Query Match 94.2%; Score 2413; DB 22; Length 475;
 Best Local Similarity 95.6%; Pred. No. 1.1e-142;
 Matches 473; Conservative 0; Mismatches 2; Indels 20; Gaps 1;
 QY 1 MTPQTOSPPFLILLITVLTVTGSGHASTPGGKERTSATORSVPSSTENKNAVSMTSSV 60
 DB 1 MTPQTOSPPFLILLITVLTVTGSGHASTPGGKERTSATORSVPSSTENKNAVSMTSSV 60
 QY 61 LSHSPGSGSSTTQGDVTLAPATEPASGSAATWGDVTSVPTTRPALGSTTPPADVTS 120
 DB 61 LSHSPGSGSSTTQGDVTLAPATEPASGSAATWGDVTSVPTTRPALGSTTPPADVTS 120
 QY 121 APDKAPAPSTAPPAHGVTSAPTRPPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180
 DB 121 APDKAPAPSTAPPAHGVTSAPTRPPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180

```
Db 121 APDNKPAFGSTAP-----AHGVTSAPTDRPAFGSTAPPAHGVT 160
Oy 181 APDNRPALGSTAPVPHNVTASGASGASASTLVHNGTSARATTPPAKSTPPSPISHS 240
Db 161 APDNRPALGSTAPVPHNVTASGASGASASTLVHNGTSARATTPPAKSTPPSPISHS 220
Oy 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSSTPDLSTGVSPFPLSHISNLOFNSSLED 300
Db 221 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSSTPDLSTGVSPFPLSHISNLOFNSSLED 280
Oy 301 PSTDYVOELORDISEMFLQIYKOGFLGSLNKKFRPGSVVQULTLAFREGTINVDVETQ 360
Db 281 PSTDYVOELORDISEMFLQIYKOGFLGSLNKKFRPGSVVQULTLAFREGTINVDVETQ 340
Oy 361 FNQKTEAASRYNLITSDVSVSHVPPFSAOSGAGVPGMGIALLVLCVLAIVYLIA 420
Db 341 FNQKTEAASRYNLITSDVSVSHVPPFSAOSGAGVPGMGIALLVLCVLAIVYLIA 400
Oy 421 LAVQCRRKNYGOLDIPAPADTYHMSSEYPTHTGRVYPPSSSTRSPYEKYSAGNGSS 480
Db 401 LAVQCRRKNYGOLDIPAPADTYHMSSEYPTHTGRVYPPSSSTRSPYEKYSAGNGSS 460
Oy 481 LSTYNPAVAATSANL 495
Db 461 LSTYNPAVAATSANL 475

RESULT 3
ABR77476
ID ABR77476 standard; Protein: 475 AA.
AC ABR77476;
DT 22-JUL-2002 (first entry)
XX
DE Human MUC1 SEQ ID NO 3.
XX
KW Human; mucin 1; MUC1; transmembrane protein; SNP; cancer; cytostatic;
KW single nucleotide polymorphism; haplotyping; genotyping; drug;
KW antiinflammatory.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FT Misc-difference 337 /note= "SNP allelic variation results in Val substituted
FT by Met"
FT
XX
XX WO200226765-A2.
XX
XX 04-APR-2002.
XX
XX 25-SEP-2001; 2001WO-US30151.
XX
XX 28-SEP-2000; 2000US-236113P.
XX
XX (GENA-) GENMISANCE PHARM INC.
XX
XX Chew A, Koshy B;
XX
XX WPI; 2002-402045/43.
XX
XX N-PSDB; ABL60158, ABL60159.
XX
XX New genetic variants of mucin 1, Transmembrane gene, useful in studying
XX expression and function of protein encoded by the gene and for
XX screening drugs to treat diseases e.g. cancer
XX
XX Claim 27; Fig 3; 75pp; English.
XX
XX The invention relates to a polynucleotide (ABL60158, ABL60159) encoding
XX mucin 1/MUC1 (ABR77476), Transmembrane isogene. The invention describes
XX novel genetic variants of the MUC1 gene. The invention is useful for
XX haplotyping/genotyping the MUC1 gene in an individual and identifying an
```

```
CC association between a trait and at least one of the haplotypes or
CC haplotype pairs of MUC1 gene. MUC1 is useful for studying the expression
CC and function of MUC1 and expressing MUC1 protein for use in screening
CC for candidate drugs to treat diseases related to MUC1 activity and in
CC studying the effect of the variation on the biological activity of MUC1
CC as well as on the binding affinity of candidate drugs targeting MUC1 for
CC the treatment of e.g. cancer. MUC1 is further used by the pharmaceutical
CC research scientist to validate MUC1 as a candidate target for and in
CC design of clinical trials of candidate drugs for treating a specific
CC condition or disease predicted to be associated with MUC1 activity.
CC MUC1 antibodies are useful in a variety of diagnostic and prognostic
CC formats and therapeutic methods.
XX
XX Sequence 475 AA;
XX
XX Query Match 94.2%; Score 2413; DB 23; Length 475;
XX Best Local Similarity 95.6%; Pred. No. 1.1e-142;
XX Matches 473; Conservative 0; Mismatches 2; Indels 20; Gaps 1;
Oy 1 MTPGCGSPFLILLITVTLYVTGSGHASTPGCEKETSATQASVSSSTERKAVEMTSV 60
Db 1 MTPGCGSPFLILLITVTLYVTGSGHASTPGCEKETSATQASVSSSTERKAVEMTSV 60
Oy 61 LSSHSPGSGSSTTGQDVTTLAPATEPASGAATWQDVTSPVTRPALCSTTPPAHDVTS 120
Db 61 LSSHSPGSGSSTTGQDVTTLAPATEPASGAATWQDVTSPVTRPALCSTTPPAHDVTS 120
Oy 121 APDNKPAFGSTAPPAHGVTSAPTDRPPGSTAPAAHGVTSAPTDRPAFGSTAPPAHGVT 180
Db 121 APDNKPAFGSTAPPAHGVTSAPTDRPPGSTAPAAHGVTSAPTDRPAFGSTAPPAHGVT 160
Oy 181 APDNRPALGSTAPVPHNVTASGASGASASTLVHNGTSARATTPPAKSTPPSPISHS 240
Db 161 APDNRPALGSTAPVPHNVTASGASGASASTLVHNGTSARATTPPAKSTPPSPISHS 220
Oy 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSSTPDLSTGVSPFPLSHISNLOFNSSLED 300
Db 221 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSSTPDLSTGVSPFPLSHISNLOFNSSLED 280
Oy 301 PSTDYVOELORDISEMFLQIYKOGFLGSLNKKFRPGSVVQULTLAFREGTINVDVETQ 360
Db 281 PSTDYVOELORDISEMFLQIYKOGFLGSLNKKFRPGSVVQULTLAFREGTINVDVETQ 340
Oy 361 FNQKTEAASRYNLITSDVSVSHVPPFSAOSGAGVPGMGIALLVLCVLAIVYLIA 420
Db 341 FNQKTEAASRYNLITSDVSVSHVPPFSAOSGAGVPGMGIALLVLCVLAIVYLIA 400
Oy 421 LAVQCRRKNYGOLDIPAPADTYHMSSEYPTHTGRVYPPSSSTRSPYEKYSAGNGSS 480
Db 401 LAVQCRRKNYGOLDIPAPADTYHMSSEYPTHTGRVYPPSSSTRSPYEKYSAGNGSS 460
Oy 481 LSTYNPAVAATSANL 495
Db 461 LSTYNPAVAATSANL 475

RESULT 4
AAB09508
ID AAB09508 standard; Protein: 473 AA.
AC AAB09508;
DT 19-NOV-2001 (first entry)
XX
XX Human mucin-1 (MUC-1) protein.
XX
XX Human milk fat globule membrane antigen; HMFg; immunostimulant;
XX cytostatic; cell mediated immune response; carcinoma; adenocarcinoma;
XX breast cancer; dendritic cell; vaccine; gene therapy; mucin-1; MUC-1;
XX immunogen.
XX
XX Homo sapiens.
XX
```

```

FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= Signal_peptide
FT Protein 22..473
FT /label= Mature_MUC-1_protein
FT Region 132..170
FT /label= VNTN_region
FT Domain 374..401
FT /label= Transmembrane_domain
XX MO200157068-A1.
XX
XX DN 09-AUG-2001.
XX
XX PF 01-FEB-2001; 2001MO-AU00090.
XX
XX PR 01-FEB-2000; 2000AU-0005369.
XX
XX PR 14-JUN-2000; 2000US-0593870.
XX
XX (AUST-) AUSTIN RES INST.
XX
XX PA McKenzie IFC, Pletersz GA, Apostolopoulos V;
XX
XX PI WPI; 2001-541537/60.
XX
XX DR Immunostimulant peptide, used as an anti-carcinoma vaccine, comprises a
XX
XX PT an epitope of the non-VNTR, non-leader region of a mucin -
XX
XX PS Disclosure; Fig 1; 84pp; English.
XX
XX CC The patent discloses peptide or polypeptides capable of eliciting
XX
XX CC an immune response, comprising an amino acid sequence corresponding
XX
XX CC to an epitope of the non-central portion of varying numbers of an
XX
XX CC amino acid motif (VNTR), non-leader region of a mucin. The peptides
XX
XX CC of the invention, fusion proteins comprising the peptide and conjugate
XX
XX CC compounds with carbohydrate polymers are used to induce a cell mediated
XX
XX CC immune response against mucin in the prevention or treatment of
XX
XX CC carcinoma, preferably adenocarcinoma, most preferably breast cancer.
XX
XX CC They are also used to pulse dendritic cell for in vivo transfer and
XX
XX CC use as a vaccine. They are also used in gene therapy. The present
XX
XX CC protein sequence is human milk fat globule membrane antigen (HMFG),
XX
XX CC mucin-1 (MUC-1).
XX
XX SQ Sequence 473 AA;
XX
Query Match 93.3%; Score 2391; DB 22; Length 473;
Best Local Similarity 95.2%; Pred. No. 2,5e-141;
Matches 471; Conservative 0; Mismatches 2; Indels 22; Gaps 2;
QY 1 MPFGQSPFFLLLLTVLTVNGSGHASTPGKETSATQSSVPSSTERNAMSSSY 60
DB 1 MPFGQSPFFLLLLTVLTVNGSGHASTPGKETSATQSSVPSSTERNAMSSSY 60
QY 61 LSHSPGSGSSSTGGQDVTATAPTEPAGSAAATMGQDVTSVYTPRALGSTTPPAHDVTS 120
DB 61 LSHSPGSGSSSTGGQDVTATAPTEPAGSAAATMGQDVTSVYTPRALGSTTPPAHDVTS 118
QY 121 APVNRKAPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
DB 119 APDNRKAPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 158
QY 181 APDNRKAPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
DB 159 APDNRKAPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 218
QY 241 TPTTLASSTGTDASTTHSHSTVPLTSSNHSSTPQLSATGVSPPFLSPHISMLQNSSLSD 300
DB 219 TPTTLASSTGTDASTTHSHSTVPLTSSNHSSTPQLSATGVSPPFLSPHISMLQNSSLSD 278
QY 301 PSTDYVOELQNDISEMFLQYKGGGFLGSLNKRPSVYVQUTAPAREGTVNHQVETQ 360
DB 279 PSTDYVOELQNDISEMFLQYKGGGFLGSLNKRPSVYVQUTAPAREGTVNHQVETQ 338

```

```

QY 361 PMQYTEAASRYNLTISDVSVHPPEPSAOSGAGVPGKGTALVLCVYALAIYYLTA 420
DB 339 PMQYTEAASRYNLTISDVSVHPPEPSAOSGAGVPGKGTALVLCVYALAIYYLTA 398
QY 421 LAVCCGRKXNYGQULDFPAPDTHPMSEVPYTHAGRYPPSSDPSPEYKSYAGNGSS 480
DB 399 LAVCCGRKXNYGQULDFPAPDTHPMSEVPYTHAGRYPPSSDPSPEYKSYAGNGSS 458
QY 481 LSTYTPAVATSAATL 495
DB 459 LSTYTPAVATSAATL 473
RESULT 5
AAAY1024
ID AAAY1024 standard; Protein; 455 AA.
XX
XX AC AAAY1024;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX DE Human Mucin 1 (MUC-1) protein fragment #5.
XX
XX KW Human; Mucin 1; MUC-1; tumour; DMR30 expression vector;
XX
XX OS anti-tumour; therapy; immune response; cytostatic; vaccine.
XX
XX OS Homo sapiens.
XX
XX PN WO200025827-A2.
XX
XX PD 11-MAY-2000.
XX
XX PF 18-OCT-1999; 99MO-EP07874.
XX
XX PR 30-OCT-1998; 98IT-MI02330.
XX
XX (MENA) MENARINI RICERCHE SPA.
XX
XX PI Parente D, Di Massimo AM, De Santis R;
XX
XX DR WPI; 2000-365410/31.
XX
XX DR N-PSDB; AAD00388.
XX
XX PT Composition containing one or more DNA molecules encoding fragments of
XX
XX PT a Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in
XX
XX PT anti-tumor therapy -
XX
XX PS Claim 16; Fig 5; 56pp; English.
XX
XX CC The present sequence is a fragment of human Mucin 1 (MUC-1), an
XX
XX CC antigenic protein overexpressed in tumour cells. The sequence was
XX
XX CC obtained from B120 tumour cells. The corresponding DNA sequence was
XX
XX CC cloned into a pMR330 expression vector and used in pharmaceutical
XX
XX CC composition e.g. vaccine for inducing an antigen-specific anti-tumour
XX
XX CC immune response. Composition containing this DNA molecule is useful in
XX
XX CC anti-tumour therapy of patients affected with tumours characterised
XX
XX CC by high MUC-1 expression.
XX
XX SQ Sequence 455 AA;
XX
Query Match 90.2%; Score 2312; DB 21; Length 455;
Best Local Similarity 95.2%; Pred. No. 2e-136;
Matches 452; Conservative 1; Mismatches 2; Indels 20; Gaps 1;
QY 21 VNGSGHASTPGKETSATQSSVPSSTERNAMSSSYLSSHPSSGSSSTGGQDVTL 80
DB 1 MNGSGHASTPGKETSATQSSVPSSTERNAMSSSYLSSHPSSGSSSTGGQDVTL 60
QY 81 AATTPAGSAAATMGQDVTSVYTPRALGSTTPPAHDVTSAPDNKXVAPGSTAPPAHGVTS 140
DB 61 AATTPAGSAAATMGQDVTSVYTPRALGSTTPPAHDVTSAPDNKXVAPGSTAPPAHGVTS 114
QY 141 APDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 200

```

```

Db 115 -----AAGTSAADINPAAGSTAPPAAGVTSAPDRKPAAGSTAPPAVNTS 160
Qy 201 AGSGASGASATLVNNTSARATTPPAKSTPPSPGIPENHSOPPTTLASHSTKTDASTHS 260
Db 161 AGSGASGASATLVNNTSARATTPPAKSTPPSPGIPENHSOPPTTLASHSTKTDASTHS 220
Qy 261 TTPPTSSNHSSTFQSTGVSPFPLSPHISNLOPNSLSLEDPSTDYQELORDISEMPLQI 320
Db 221 TTPPTSSNHSSTFQSTGVSPFPLSPHISNLOPNSLSLEDPSTDYQELORDISEMPLQI 280
Qy 321 YKQSGFLGLSNIKRPGSVVOLLAFREGTINVDVETQFNQKTEAARVNLISDVS 380
Db 281 YKQSGFLGLSNIKRPGSVVOLLAFREGTINVDVETQFNQKTEAARVNLISDVS 340
Qy 381 VSHVFPFSSAGAGVPGMGIALVLVLCVLAIVYLLALAVCCRRKXVYQGLDIFPAR 440
Db 341 VSDVFPFSSAGAGVPGMGIALVLVLCVLAIVYLLALAVCCRRKXVYQGLDIFPAR 400
Qy 441 DTYHFMSEYPTTHTGRVYPPSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 495
Db 401 DTYHFMSEYPTTHTGRVYPPSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 455

RESULT 6
AAV71030
ID AAV71030 standard; Protein; 577 AA.
AC AAV71030;
XX
XX 29-AUG-2000 (first entry)
DE Ubiqutin-B. coli lacti-human Mucin 1 fusion protein #5.
XX
XX Ubiqutin; Lacti, beta-galactosidase; fusion protein;
KM human; Mucin 1; MUC-1; tumour; pMR330 expression vector;
XX anti-tumour; therapy; immune response; cytostatic; vaccine.
OS Chimeric - Homo sapiens.
OS Chimeric - Escherichia coli.
XX
FH Key Location/Qualifiers
FT 1..123 /label= UbiLacti protein
FT /note= "contains ubi-ubiquitin sequence fused to a
FT portion of B. coli Lacti"
FT 124..577 /label= Human_MUC-1_fragment
FT Region
XX
XX W0200025827-A2.
XX
XX 11-MAY-2000.
XX
XX 18-OCT-1999; 99MO-EP07874.
XX
XX 30-OCT-1998; 98IT-MI02330.
XX
XX (MENA ) MENARINI RICERCHE SPA.
XX
XX Parente D, Di Maseimo AM, De Santis R;
XX
XX WPI; 2000-365410/31.
XX
XX N-PSDB; AAD00394.
XX
XX Composition containing one or more DNA molecules encoding fragments of
XX a Mucin 1 (MUC-1) protein overexpressed in tumor cells; useful in
XX anti-tumor therapy -
XX
XX Claim 18; Fig 11; 56pp; English.
XX
XX The present sequence is a fusion protein consisting
XX of human Mucin 1 (MUC-1) fragment fused to UbiLacti sequence at the
XX N-terminus. The UbiLacti sequence consists of ubi-ubiquitin from MCF7 cell

```

```

CC line and a portion of B. coli beta-galactosidase (lacti). MUC-1 is an
CC antigenic protein overexpressed in tumour cells. The corresponding
CC DNA sequence is cloned into a pMR330 expression vector and used in
CC pharmaceutical composition e.g. vaccine for inducing an antigen-specific
CC anti-tumour immune response. Composition containing this DNA molecule
CC is useful in anti-tumour therapy of patients affected with tumours
CC characterised by high MUC-1 expression.
XX
XX Sequence 577 AA;
SQ
Query Match 90.2%; Score 2311.5; DB 21; Length 577;
Best Local Similarity 91.4%; Pred. No. 2.9e-136;
Matches 459; Conservative 2; Mismatches 14; Indels 27; Gaps 3;
Qy 1 MTEGTO--SP-----FFLLLLLTLLTVVYSGHASTPGEKEKTSATQSSVPSSTEXNA 53
Db 96 LAENTQTASPRALADSLMQLARQVSRGSGHASTPGEKEKTSATQSSVPSSTEXNA 155
Qy 54 VSMTSSVLSHSPGSGSSTTQGDVTLAPATEPAGSAAFWGQDVTSVPVTRPALGSTTP 113
Db 156 VSMTSSVLSHSPGSGSSTTQGDVTLAPATEPAGSAAFWGQDVTSVPVTRPALGSTTP 215
Qy 114 PAADVTSAPDNKPAAGSTAPPAHGYTSAPDTRPPGSGTAPAAHGYTSAPDTRPPAGSTAP 173
Db 216 PAADVTSAPDNKPAAGSTAP-----AHGYTSAPDTRPPAGSTAP 255
Qy 174 PAHGYTSAPDNKPAAGSTAPPAHNTVSASGSASGASTLVHNGTSARATTPPAKSTPPS 233
Db 256 PAHGYTSAPDNKPAAGSTAPPAHNTVSASGSASGASTLVHNGTSARATTPPAKSTPPS 315
Qy 234 IPSHSDPTTTLASHSTKTDASTHSTVPPLTSSNNHSTSPLSTGVSFFFLSFHISNLQ 293
Db 316 IPSHSDPTTTLASHSTKTDASTHSTVPPLTSSNNHSTSPLSTGVSFFFLSFHISNLQ 375
Qy 294 FNSLSLEDPSTDYQELORDISEMPLQIYKQSGFLGLSNIKRPGSVVOLLAFREGTIN 353
Db 376 FNSLSLEDPSTDYQELORDISEMPLQIYKQSGFLGLSNIKRPGSVVOLLAFREGTIN 435
Qy 354 VHVVEFQFNQKTEAARVNLITSDVSVHVPFSSAGAGVPGMGIALVLVLCVLA 413
Db 436 VHVVEFQFNQKTEAARVNLITSDVSVDVFPFSSAGAGVPGMGIALVLVLCVLA 495
Qy 414 AIVYLLALAVCCRRKXVYQGLDIFPARDTHPMSEYPTTHTGRVYPPSTDRSPYEKVS 473
Db 496 AIVYLLALAVCCRRKXVYQGLDIFPARDTHPMSEYPTTHTGRVYPPSTDRSPYEKVS 555
Qy 474 AGNGSSLSYTNPAVAATSANL 495
Db 556 AGNGSSLSYTNPAVAATSANL 577

RESULT 7
AAR23973
ID AAR23973 standard; Protein; 455 AA.
AC AAR23973;
XX
XX 06-NOV-1992 (first entry)
XX
XX Transmembrane form of H23-ETA antigen.
XX
XX ETA-T; human epithelial antigen; Monoclonal antibody H23; vaccine;
XX malignant tumour; breast cancer; tandem repeat.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Peptide 1..21 /label= signal
FT Protein 22..455 /label= ETA-T
FT Misc-difference 134 /label= Pro, Ala
FT

```

FT	Misc-difference	144	/note= "natural polymorphism"
FT		/label= Thr, Asn	
FT	Misc-difference	147	/note= "natural polymorphism"
FT		/label= Pro, Ala	
XX		/note= "natural polymorphism"	
PN	N09207000-A.		
PD	30-APR-1992.		
DP	23-OCT-1991;	9IMC-FR00835.	
PK	23-OCT-1990;	90FR-0013101.	
XX	(TRGE) TRANSGENE SA.		
PA	Chambon P, Harevenni M, Kieny MP, Latche R;		
XX	WPI; 1992-167097/20.		
DR	N-PsDB; AAQ23276.		
PT	Compens. cong. polypeptide antigen recognised by antibody H23 -		
PT	for treatment of mammary tumours, also for pox virus compens. for		
PT	use in vaccines		
PS	Claim 3; Page 15-18; 29pp; French.		
XX	The tumour antigen recognised by antibody H23 is aberrantly		
CC	expressed in epithelial cells from cancerous mammary tissue in about		
CC	90 per cent of breast cancer cases; in a normal individual		
CC	expression is negligible. The antigen exists in two forms:		
CC	transmembrane (ETA-T) and secreted (ETA-S). Both forms show a high		
CC	degree of polymorphism. A 20 amino acid subunit in ETA can be		
CC	randomly repeated up to 80 times. From one subunit to the next, 1 to		
CC	3 amino acids can differ. DNA coding for immunogenic fragments of		
CC	ETA can be inserted into e.g. vaccinia viruses for treatment of		
CC	mammary tumours		
CC	See also AAQ24678-Q24681, AAQ23277 and AAQ23974-R23981.		
SQ	Sequence	455 AA;	
Query Match	89.6%; Score 2295; DB 13; Length 455;		
Best Local Similarity	91.3%; Pred. No. 2,3e-135;		
Matches 452; Conservative	0; Mismatches 3; Indels 40; Gaps 1;		
OY	1 MTPGTQSPEFLILLITLVTVTGGHASTPGEKEKTSATQRSSVSTERNKAVSMTSY 60		
DB	1 MTPGTQSPEFLILLITLVTVTGGHASTPGEKEKTSATQRSVSSTERNKAVSMTSY 60		
OY	61 LSLSHPGSGSSTTGODVTTLAEPAGSAATWQDVTSVPVTRPALGSTTPPAHDVTS 120		
DB	61 LSLSHPGSGSSTTGODVTTLAEPAGSAATWQDVTSVPVTRPALGSTTPPAHDVTS 120		
OY	121 APDNKPACGSTAPPAGGVTSADPTRPPPGSTAPAAHGVTSAFDTRPAPGSTAPPAHVTS 180		
DB	121 APDNKPACGSTAPPAGGVTSADPTRPPPGSTAPAAHGVTSAFDTRPAPGSTAPPAHVTS 180		
OY	181 APDNKPACGSTAPPAGGVTSADPTRPPPGSTAPAAHGVTSAFDTRPAPGSTAPPAHVTS 240		
DB	145 ----RPXGISTAPPVNVNTSAGSGASASTLVHNGTSARATTTTPASKSTPPISPHSD 200		
OY	241 TPTTLASHSTSTDASSSTHSSTVPPLTSSNHNSTSPOLSTGVFFFLFPHSNLOFNSSLLE 300		
DB	201 TPTTLASHSTSTDASSSTHSSTVPPLTSSNHNSTSPOLSTGVFFFLFPHSNLOFNSSLLE 260		
OY	301 PSTDYVOGLGDISEMFLDTKYGGGFLGNLTFPRGSIVVOLTLAFRGSTINADVETD 360		
DB	261 PSTDYVOGLGDISEMFLDTKYGGGFLGNLTFPRGSIVVOLTLAFRGSTINADVETD 320		
OY	361 FNYDKTEAASNNLTISDVSVSHVPPFPSSAQSGAGVPGWGIALILVVCULVALATVILA 420		
DB	321 FNYDKTEAASNNLTISDVSVSHVPPFPSSAQSGAGVPGWGIALILVVCULVALATVILA 380		
OY	421 LAVOCRRKNVGOLDIFPADDTYMNSGYPRHYTHORVYPSPSTDSPEFYKSANCGSS 480		
DB	381 LAVOCRRKNVGOLDIFPADDTYMNSGYPRHYTHORVYPSPSTDSPEFYKSANCGSS 440		
OY	481 LSTYNPAVAARSANL 495		
DB	LSTYNPAVAARSANL 455		
RESULT 8			
AAM77233	ID	AAM77233 standard; Protein; 508 AA.	
XX	AAW77233;		
AC	20-NOV-1998 (first entry)		
DT	MiniMUC1 protein.		
DE	MUC1; recombinant pox virus; cytotoxic T-lymphocyte; immunogen;		
KW	tumour; tumour-associated antigen.		
XO	Homo sapiens.		
OS	WO9837095-A2.		
PN	27-AUG-1998.		
PD	24-FEB-1996; 98MO-US03693.		
PF	24-FEB-1997; 97US-0038253.		
PR	(DAND) DANA PARBER CANCER INST INC.		
PA	(THER-) THERION BIOLOGICS CORP.		
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.		
XX	Gritz L, Kantor J, Kufe D, Panicali D, Schlom J;		
PI	WPI; 1998-467492/40.		
DR	N-PsDB; AAU48329.		
XX	New recombinant pox virus for tumour therapy - comprises DNA		
PT	encoding an immunogenic mini-MUC1 fragment comprising 5-25 MUC1		
PT	tandem repeat units		
PS	Example 1; Page 21-22; 42pp; English.		
XX	The immunogenic mini-MUC1 fragment was created from MUC1 tandem repeat		
CC	units for inclusion in a recombinant pox virus (BPV). The BPV was used		
CC	in a pharmaceutical composition also containing an immunomodulator to		
CC	generate MUC1 specific cytotoxic T-lymphocytes. The recombinant pox		
CC	virus therefore encodes an immunogenic MUC1 fragment that does not		
CC	undergo significant genetic deletion, thereby providing an unexpectedly		
CC	stable and immunogenic pox virus. They can be used to prevent or treat		
CC	tumours expressing MUC1 tumour-associated antigens.		
SQ	Sequence	508 AA;	
Query Match	89.1%; Score 2282.5; DB 19; Length 508;		
Best Local Similarity	91.2%; Pred. No. 1.6e-134;		
Matches 455; Conservative	6; Mismatches 32; Indels 17; Gaps 3;		
OY	1 MTPGTQSPEFLILLITLV-----TVTVGSHASTPGEKEKTSATQRSSVSSTERNK 51		
DB	1 MTPGTQSPEFLILLITLVTKTPAPVTVTGGHASTPGEKEKTSATQRSSVSSTERNK 60		
OY	52 NAYSMTSIVLSHSPGSGSSTTGODVTTLAEPAGSAATWQDVTSVPVTRPALGST 111		
DB	61 NAYSMTS--LINSGVGRGSTAPPAGGVTSADPTRPPPGSTAPAAHGVTSAFDTRPAPGST 118		
OY	112 TPRAHDVTSAPDNKPACGSTAPPAGGVTSADPTR-----PPPGSTAPAAHGVTSAFDTR 165		

PN JP07051065-A.
 XX
 PD 28-FEB-1995.
 XX
 PF 21-FEB-1992; 92JP-0035085.
 XX
 PR 21-FEB-1992; 92JP-0035085.
 XX
 PA (NIKO-) NIPPON KOTAI KENKUSHO KK.
 PA (UYKA-) UNIV KAGOSHIMA.
 XX
 DR WPI; 1995-127356/17.
 DR N-PSDB; AAT29190.
 XX
 PT Glyco:protein 39 gene - used in the mass production of glyco:protein
 PT 39, for use as tumour and immune abnormality marker
 XX
 PS Claim 2; Page 9-10; 10pp; Japanese.
 CC The sequences given in AAR96297-98 represent portions of glycoprotein
 CC 39. This sequence represents the C terminal portion of the glycoprotein
 CC 39 gene. The cDNA encoding this sequence was isolated from lambda gt11
 CC cDNA library derived from the gastric cancer cell line KATO-III
 CC Glycoprotein 39 is expected to be used as a tumour marker, an immune
 CC abnormality marker or a marker for various inflammatory diseases.
 XX
 SQ Sequence 327 AA;
 Query Match 65 28; Score 1671; DB 16; Length 327;
 Best Local Similarity 99.1%; Pred. No. 1.3e-96;
 Matches 324; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 169 GSTPPAPAGTAPNRPALGSTAPVHNVTASGASGASTLVNNGTSARATTPPASK 228
 DB 1 GSTPPAPAGTAPNRPALGSTAPVHNVTASGASGASTLVNNGTSARATTPPASK 60
 QY 229 STPTSTSHSTPTTLASHTKTDASTHSTVPELTSSNHSFPOLSTGVSPFTLSR 288
 DB 61 STPTSTSHSTPTTLASHTKTDASTHSTVPELTSSNHSFPOLSTGVSPFTLSR 120
 QY 289 ISNQRNSLEPSTDYQELQNDISEMFLQYKGGFGLSNIKRPGSVVQGLTAER 348
 DB 121 ISNQRNSLEPSTDYQELQNDISEMFLQYKGGFGLSNIKRPGSVVQGLTAER 180
 QY 349 EGTINHDVETOPNOYKTEASRNLTISDVSVSHVPPPSAQSAGVPGMGLALVLC 408
 DB 181 EGTINHDVETOPNOYKTEASRNLTISDVSVSHVPPPSAQSAGVPGMGLALVLC 240
 QY 409 VLVALAIVYLALAVCCRRKNTGOLDIFPARDTYHMSERYPTTHGRVYVPSSTRSP 468
 DB 241 VLVALAIVYLALAVCCRRKNTGOLDIFPARDTYHMSERYPTTHGRVYVPSSTRSP 300
 QY 469 YEKVSAGNGSSLSYTNPAVAATSAVL 495
 DB 301 YEKVSAGNGSSLSYTNPAVAATSAVL 327
 RESULT 11
 AAB43416
 ID AAB43416 standard; Protein; 321 AA.
 AC AAB43416;
 XX
 DT 08-FEB-2001 (first entry)
 DE Human cancer associated protein sequence SEQ ID NO:861.
 XX
 KW Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
 KW antidiabetic; antiaesthetic; antithematic; antiaesthetic; antiviral;
 KW antiinflammatory; antihydroid; antiallergic; antibacterial; cardiac;
 KW dermatologic; neuroprotective; thrombolytic; coagulant; nociceptive;
 KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;

KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening.
 OS Homo sapiens.
 PN WO20005350-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05882.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-587533/55.
 DR N-PSDB; AAC78625.
 XX
 PT Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer -
 XX
 PS Claim 11; Page 1414-1416; 2352pp; English.
 CC AAC78607 to AAC78448 encode the human cancer associated proteins given
 CC -in AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnery; immunomodulator;
 CC antidiabetic; antiaesthetic; antithematic; antiaesthetic; antiviral;
 CC antiinflammatory; antihydroid; antiallergic; antibacterial; cardiac;
 CC dermatologic; neuroprotective; thrombolytic; coagulant;
 CC nociceptive; vasotropic; antipsoriatic; antiangiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 321 AA;
 Query Match 64.0%; Score 1640; DB 21; Length 321;
 Best Local Similarity 99.1%; Pred. No. 1.1e-94;
 Matches 318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 175 AHGVTAPNRPALGSTAPVHNVTASGASGASTLVNNGTSARATTPPASKTPPSI 234
 DB 1 AHGVTAPNRPALGSTAPVHNVTASGASGASTLVNNGTSARATTPPASKTPPSI 60
 QY 235 PSHNSDPTTLASHTKTDASTHSTVPELTSSNHSFPOLSTGVSPFTLSR 294
 DB 61 PSHNSDPTTLASHTKTDASTHSTVPELTSSNHSFPOLSTGVSPFTLSR 120
 QY 295 NSLSEPDSTDYQELQNDISEMFLQYKGGFGLSNIKRPGSVVQGLTAEREGTIN 354
 DB 121 NSLSEPDSTDYQELQNDISEMFLQYKGGFGLSNIKRPGSVVQGLTAEREGTIN 180
 QY 355 HDVETOPNOYKTEASRNLTISDVSVSHVPPPSAQSAGVPGMGLALVLCVVALA 414
 DB 181 HDVETOPNOYKTEASRNLTISDVSVSHVPPPSAQSAGVPGMGLALVLCVVALA 240
 QY 415 IVYLALAVCCRRKNTGOLDIFPARDTYHMSERYPTTHGRVYVPSSTRSPYEKVA 474

DB 241 IYVLIAMAVCQCRKNYGGOLDIFPARDTYHMSSEYPTHTHGRVPPSTDSPEYKVA 300
 QY 475 GNGSSISYTNPAVAATSA 495
 DB 301 GNGSSISYTNPAVAATSA 321

RESULT 12
 ID AA084810 standard; Protein: 312 AA.
 AC AA084810;
 DT 08-MAY-2002 (first entry)
 DE Human MUC1R consensus sequence.
 KW Saviine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
 KW viral infection; human immunodeficiency virus; melanoma;
 KW bacterial infection; Salmonella; Legionella; parasitic infection;
 KW Typhosoma; Toxoplasma; Giardia.
 OS Homo sapiens.
 PN WO200190197-A1.
 PD 29-NOV-2001.
 PP 25-MAY-2001; 2001WO-AU00622.
 PR 26-MAY-2000; 2000AU-000761.
 XX 26-MAY-2000; 2000AU-000761.
 PA (AUSU) UNIT AUSTRALIAN NAT.
 PI Thomson SA, Ramshaw JA;
 DR WPI: 2002-147575/19.

XX New synthetic polypeptides having several different segments of at
 PT least one parent polypeptide linked together differently compared to
 PT the linkage in the parent polypeptide, for inducing immune response
 PT against a pathogen or cancer -
 XX
 PS Example 3; Fig 27; 364pp; English.

XX The invention relates to a new synthetic polypeptide (1) comprising
 CC several different segments of at least one parent polypeptide linked
 CC together in a different relationship relative to their linkage in the
 CC parent polypeptide to impede, abrogate or otherwise alter at least one
 CC function associated with the parent polypeptide and for inducing an
 CC immune response against a pathogen or cancer. Also included are a
 CC synthetic polynucleotide encoding and a computer system for
 CC designing the synthetic polypeptides. The synthetic polypeptides and
 CC polynucleotides are referred to as a Saviine. The synthetic polypeptide is
 CC useful for modulating immune responses preferably directed against a
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, bone
 CC colon, head and neck, pancreas, prostate, stomach, kidney, bone
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present
 CC sequence is a consensus sequence for a parent protein used to design a
 CC Saviine of the invention.

XX Sequence 312 AA:
 Query Match 61.9%; Score 1587; DB 23; Length 312;
 Best Local Similarity 98.7%; Pred. No. 2.1e-91;

Matches 308; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 164 NRPLGATAPVNAVNTSAGSAGSASTLVHNGTSARATTPPASKSPFPIPSHSDPT 243
 DB 1 NRPALGATAPVNAVNTSAGSAGSAGSASTLVHNGTSARATTPPASKSPFPIPSHSDPT 60
 QY 244 TLASHSTKTDASTHSTHSTVPPPLTSNNSTSPOLSTGVSPFPLSHINLOFNSLADPT 303
 DB 61 TLASHSTKTDASTHSTHSTVPPPLTSNNSTSPOLSTGVSPFPLSHINLOFNSLADPT 120
 QY 304 DYQELQORDISENFLOIYKQGFGLSNIKFRPGSVVQVLTAFREGTINVHDVETQNO 363
 DB 121 DYQELQORDISENFLOIYKQGFGLSNIKFRPGSVVQVLTAFREGTINVHDVETQNO 180
 QY 364 YKTEAASRYNLTISDVSVSHVPPPSAQSAGVPGWGLALVLCVLAVALIYVLIAMAV 423
 DB 181 YKTEAASRYNLTISDVSVSHVPPPSAQSAGVPGWGLALVLCVLAVALIYVLIAMAV 240
 QY 424 CQCRKNYGGOLDIFPARDTYHMSSEYPTHTHGRVPPSTDSPEYKVA 483
 DB 241 CQCRKNYGGOLDIFPARDTYHMSSEYPTHTHGRVPPSTDSPEYKVA 300

QY 484 TNPVAATSA 495
 DB 301 TNPVAATSA 312

RESULT 13
 ID AAR27665 standard; Protein: 287 AA.
 AC AAR27665;
 DT 06-NOV-1992 (first entry)
 DE Secreted form of H23-ETA antigen.
 KW ETA-S; human epithelial antigen; Monoclonal antibody H23; vaccine;
 KW malignant tumour; breast cancer; tandem repeat.
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..21
 FT /label= signal
 FT Protein 22..287
 FT /label= ETA-T
 FT Misc-difference 134
 FT /label= Pro, Ala
 FT /note= "natural polymorphism"
 FT Misc-difference 144
 FT /label= Thr, Asn
 FT /note= "natural polymorphism"
 FT Misc-difference 147
 FT /label= Pro, Ala
 FT /note= "natural polymorphism"

XX WO9207000-A.
 XX 30-APR-1992.
 XX 23-OCT-1991; 91WO-FR00835.
 XX 23-OCT-1990; 90FR-0013101.
 XX (TRGE) TRANSGENE SA.
 XX Chandon P, Hareuveni M, Kieny MP, Lathe R;
 DR WPI: 1992-167097/20.
 DR N-PSDB; AAQ29277.
 PT Compsns. contg. polypeptide antigen recognised by antibody H23 -

PT for treatment of mammary tumours, also for pox virus compans. for
 XX use in vaccines

PS Claim 3; Page 19-21; 29pp; French.

XX The tumour antigen recognised by antibody H23 is aberrantly
 CC expressed in epithelial cells from cancerous mammary tissue in about
 CC 90 per cent of breast cancer cases; in a normal individual
 CC expression is negligible. The antigen exists in two forms:
 CC transmembrane (ETA-T) and secreted (ETA-S). Both forms show a high
 CC degree of polymorphism. A 20 amino acid subunit in ETA can be
 CC tandemly repeated up to 80 times. From one subunit to the next, 1 to
 CC 3 amino acids can differ. DNA coding for immunogenic fragments of
 CC ETA can be inserted into e.g. vaccinia viruses for treatment of
 CC mammary tumours.
 CC See also AAQ24678-Q24681, AAQ29276-7 and AAR23974-R23981.

XX Sequence 287 AA;

Query Match 52.8%; Score 1352; DB 13; Length 287;
 Best Local Similarity 86.1%; Pred. No. 8.7e-77;
 Matches 272; Conservative 0; Mismatches 4; Indels 40; Gaps 1;

QY 1 MTPGQSPFFLLLLTLVTVVSGHASTPGGKETSATORSVPSSTERNVMTSSV 60
 DB 1 MTPGQSPFFLLLLTLVTVVSGHASTPGGKETSATORSVPSSTERNVMTSSV 60
 QY 61 LSHSPGSGSSTTGQDVTLPATPAPASGSAATWGQDVTSPVTPPALGSTPPADHVT 120
 DB 61 LSHSPGSGSSTTGQDVTLPATPAPASGSAATWGQDVTSPVTPPALGSTPPADHVT 120
 QY 121 APDNKPAAGSTAPPAHGVTSAPDTPPPGSTARPAHGVTSAPDTPAPASGTAAPAHGVTS 180
 DB 121 APDNKPAAGSTAPPAHGVTSAPDTPPPGSTARPAHGVTSAPDTPAPASGTAAPAHGVTS 180
 QY 181 APDNKPAAGSTAPPAHGVTSAPDTPPPGSTARPAHGVTSAPDTPAPASGTAAPAHGVTS 240
 DB 145 ---PPXGSTARPAHGVTSAPDTPPPGSTARPAHGVTSAPDTPAPASGTAAPAHGVTS 200
 QY 241 TPTTLASHSTKTDASTHSTVPPVLTSSNHSSTPOLSTGVSPFPLSPHISNLOPNSLSD 300
 DB 201 TPTTLASHSTKTDASTHSTVPPVLTSSNHSSTPOLSTGVSPFPLSPHISNLOPNSLSD 260
 QY 301 PSTDYQELQORDISEM 316
 DB 261 PSTDYQELQORDISEM 276

RESULT 14

AAR89418 standard; Protein; 273 AA.

AC AAR89418;

DT 24-APR-1996 (first entry)

DE Mucin-derived protein MUC1/X.

XX MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
 KM MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
 KM receptor; diagnosis; imaging; therapy.

OS Homo sapiens.

XX MO9603502-A2.

PD 08-FEB-1996.

PF 21-JUL-1995; 95MO-IB00627.

PR 26-JUL-1994; 94IL-0110464.

XX (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.

XX Wreschner DH;

DR WPI; 1996-117047/12.

DR N-PSDB; AAT10677.

XX Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
 PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
 PT cancer

PS Claim 4; Fig 5A; 79pp; English.

XX Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
 CC MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly
 CC differentially expressed in human breast cancer tissue and significantly
 CC enhance the in vivo tumorigenic potential of mammary epithelial cells.
 CC They serve as cell surface receptor molecules participating in signal
 CC transduction. The proteins can be obcd. by expression of encoding cDNA
 CC (see AAT10677-82) in recombinant host cells. They are used in the
 CC treatment of human breast cancer and as diagnostic reagents. Receptor
 CC ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
 CC isolated.

XX Sequence 273 AA;

Query Match 49.7%; Score 1273; DB 17; Length 273;
 Best Local Similarity 54.9%; Pred. No. 6.9e-72;
 Matches 272; Conservative 0; Mismatches 1; Indels 222; Gaps 1;

QY 1 MTPGQSPFFLLLLTLVTVVSGHASTPGGKETSATORSVPSSTERNVMTSSV 60
 DB 1 MTPGQSPFFLLLLTLVTVVSGHASTPGGKETSATORSVPSSTERNVMTSSV 60
 QY 61 LSHSPGSGSSTTGQDVTLPATPAPASGSAATWGQDVTSPVTPPALGSTPPADHVT 120
 DB 61 LSHSPGSGSSTTGQDVTLPATPAPASGSAATWGQDVTSPVTPPALGSTPPADHVT 120
 QY 121 APDNKPAAGSTAPPAHGVTSAPDTPPPGSTARPAHGVTSAPDTPAPASGTAAPAHGVTS 180
 DB 121 APDNKPAAGSTAPPAHGVTSAPDTPPPGSTARPAHGVTSAPDTPAPASGTAAPAHGVTS 180
 QY 181 APDNKPAAGSTAPPAHGVTSAPDTPPPGSTARPAHGVTSAPDTPAPASGTAAPAHGVTS 240
 DB 181 APDNKPAAGSTAPPAHGVTSAPDTPPPGSTARPAHGVTSAPDTPAPASGTAAPAHGVTS 240
 QY 241 TPTTLASHSTKTDASTHSTVPPVLTSSNHSSTPOLSTGVSPFPLSPHISNLOPNSLSD 300
 DB 241 TPTTLASHSTKTDASTHSTVPPVLTSSNHSSTPOLSTGVSPFPLSPHISNLOPNSLSD 260
 QY 301 PSTDYQELQORDISEM 316
 DB 261 PSTDYQELQORDISEM 276

RESULT 15

AAR89419 standard; Protein; 282 AA.

AC AAR89419;

DT 24-APR-1996 (first entry)

XX Mucin-derived protein MUC1/X/alt.
 DB MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
 XX MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
 KM receptor; diagnosis; imaging; therapy.
 XX Homo sapiens.
 OS MO9603502-A2.
 XX
 XX 08-FEB-1996.
 XX PD
 XX 21-JUL-1995; 95WO-1B00627.
 XX PF
 XX 26-JUL-1994; 94IL-0110464.
 XX PR
 XX (UDRA-) UNIV RAMOT APPL RES & IND DEV LTD.
 XX PA
 XX Wrechner DH;
 XX PI
 XX MPI; 1996-117047/12.
 XX DR
 XX N-PSDB; AAT10678.
 XX
 XX Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
 PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
 PT cancer
 XX
 PS Claim 4; Fig 5B; 79pp; English.
 XX
 CC Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
 CC MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR9418-23) are highly and
 CC differentially expressed in human breast cancer tissue and significantly
 CC enhance the in vivo tumorigenic potential of mammary epithelial cells.
 CC They serve as cell surface receptor molecules participating in signal
 CC transduction. The proteins can be obtd. by expression of encoding cDNA
 CC (see AAT10677-82) in recombinant host cells. They are used in the
 CC treatment of human breast cancer and as diagnostic reagents. Receptor
 CC ligands MUC1/W, W/alt, Z and Z/alt (AAR9424-27) have also been
 CC isolated.
 CC
 XX
 XX SQ Sequence 282 AA;

Query Match 49.1%; Score 1258.5; DB 17; Length 282;
 Best Local Similarity 54.0%; Pred. No. 5,8e-71;
 Matches 272; Conservative 0; Mismatches 1; Indels 23; Gaps 2;
 QY 1 MTPGTGSPFFLLILLTL-----TVTSGSHASSTPGGKETSATORSSVPSSTEK 51
 DB 1 MTPGTGSPFFLLILLTLVLTATTAPKPAVTVTGSHASSTPGGKETSATORSSVPSSTEK 60
 QY 52 NAVSWTSSVLSHSPGSGSTTGQDVTIAPATEPASGAATWGDVTSVPVTRPALGST 111
 DB 61 NA----- 62
 QY 112 TRPAHDVTSAPDNKPAKPGSTAPPAHGVTSAPDTRPPGSTAPAAHGVTSAPDTRPAKST 171
 DB 63 ----- 62
 QY 172 APPAHGVTSAPDNKPAKPGSTAPPAHGVTSAPDTRPPGSTAPAAHGVTSAPDTRPAKST 231
 DB 63 ----- 62
 QY 233 PSIPSHSDPTPTTLASHSTKTDASSTHSTVPPILTSNHSTSPOLSTGVSPFFLSFHISN 291
 DB 63 -----LSTGVSPFFLSFHISN 78
 QY 292 LQFNSSLEDESTVYOELORDISEMFLQIYKGGFLGLSNIKFRPGSVVQTLTAPRST 351
 DB 79 LQFNSSLEDESTVYOELORDISEMFLQIYKGGFLGLSNIKFRPGSVVQTLTAPRST 138
 QY 352 INVHDEVTQFNQKTEASRYNLITSDVSVHVPFPSAQSAGVPGWGLALLVLCVLY 411

DB 139 INVHDEVTQFNQKTEASRYNLITSDVSVHVPFPSAQSAGVPGWGLALLVLCVLY 198
 QY 412 ALATVYLIALAVGCCRRKVGOLDIPAPDTHPMSEYPTHTGGRVPPSSSTRDSPTXK 471
 DB 199 ALATVYLIALAVGCCRRKVGOLDIPAPDTHPMSEYPTHTGGRVPPSSSTRDSPTXK 258
 QY 472 VSAAGGSSLSYTNPAVAATSANL 495
 DB 259 VSAAGGSSLSYTNPAVAATSANL 282

Search completed: May 1, 2003, 23:10:26
 Job time : 66.8393 secs

Page 1

GenCore version 5.1.4_p5_4578
(c) 1993 - 2003 CompuGen Ltd.

Search time 105.089 Seconds

970.540 million cell updates/sec

...NGGSSLSTNPAAATSANL 495

BLOSUM62

671580

1: sp.archaea:.*
2: sp.bacteria:.*
3: sp.fungi:.*
4: sp.human:.*
5: sp.invertebrate:.*
6: sp.mammal:.*
7: sp.mhc:.*
8: sp.organelle:.*
9: sp.phage:.*
10: sp.plant:.*
11: sp.podent:.*
12: sp.virus:.*
13: sp.veterebrate:.*
14: sp.unclassified:.*
15: sp.rvius:.*
16: sp.bacteriaph:.*
17: sp.archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Length	DB	ID	Description
1	1795.5	70.1	553	6	Q9M2L1	Q9m2l1 macca mla
2	1234	48.2	602	6	O19115	O19115 cryciolagus
3	1228*5	48.0	580	6	Q8M414	Q8m414 bos taurus
4	11398	46.8	621	11	Q93K60	Q93k60 mus musculus
5	1120.5	43.7	554	6	Q95J89	Q95j89 bos taurus
6	720.5	28.1	193	11	Q60S51	Q60s51 mesocricetus
7		21.9	192	6	Q28078	Q28078 bos taurus
8	706.5	27.6	193	6	Q28723	Q28723 cryciolagus
9	699.5	27.3	193	11	O60408	O60408 cavia cavia
10	358	14.0	74	11	O35770	O35770 rattus norv
11	267.5	10.4	487	2	Q9ZB39	Q9zb39 ureaplasma
12	265.5	10.4	1334	16	Q9RKR9	Q9rkr9 streptococ
13	264.5	10.3	169	5	Q9N4S7	Q9n4s7 caenorhabdi
14	254.5	9.9	160	6	Q8SQ37	Q8sq37 bos taurus
15	243.5	9.9	1225	5	Q9BKV7	Q9bkv7 leishmania
16	248.5	9.7	160	6	Q8SQ36	Q8sq36 bos taurus

17	244.5	9.5	1349	4	Q8M04	Q8W44 homo sapien
18	240	9.4	961	3	Q92223	Q92223 emericella
19	239	9.3	160	6	Q8M38	Q8M38 bos taurus
20	237	9.3	2232	5	P91365	P91365 caenorhabdi
21	234.5	9.2	907	12	Q66537	Q66537 homo herpex
22	234	9.1	377	4	Q8M37	Q8M37 homo sapien
23	233	9.1	3178	5	Q969D4	Q969D4 caenorhabdi
24	232.5	9.1	3570	4	Q99552	Q99552 homo sapien
25	232	9.1	786	5	Q21037	Q21037 caenorhabdi
26	232.2	9.1	866	12	Q56854	Q56854 human herpex
27	231.5	9.0	528	6	Q29071	Q29071 sus scrofa
28	230	9.0	957	4	Q9UKD	Q9UKD homo sapien
29	229	8.9	873	5	Q9Y075	Q9Y075 leishmania
30	225.5	8.8	393	5	Q9Y075	Q9Y075 leishmania
31	225	8.8	293	4	Q41898	Q41898 homo sapien
32	222	8.7	624	3	Q94317	Q94317 schizosach
33	222	8.7	352	4	Q41861	Q41861 homo sapien
34	221	8.6	866	12	Q9Q87	Q9Q87 human herpex
35	220.5	8.6	1029	4	Q8IDH7	Q8IDH7 homo sapien
36	220.5	8.6	1992	3	Q9PE11	Q9PE11 neurospora
37	220	8.6	1612	5	Q9Y92	Q9Y92 dirosophila
38	219.5	8.6	1275	5	Q76602	Q76602 caenorhabdi
39	219	8.5	317	6	Q29070	Q29070 sus scrofa
40	218.5	8.5	322	11	Q62605	Q62605 rattus norv
41	217.5	8.5	783	12	Q91331	Q91331 cercopithec
42	217	8.5	1203	5	Q9NSK0	Q9NSK0 caenorhabdi
43	216.5	8.5	851	12	Q9Q8H6	Q9Q8H6 human herpex
44	215.5	8.4	4776	6	Q97E71	Q97E71 streptococc
45	215	8.4	8766	12	Q8V0L5	Q8V0L5 equine heppe

ALIGNMENTS

RESULT 1

ID	Q9MZL1		PRELIMINARY;	PRT;	553 AA.
AC	Q9MZL1				
DT	01-OCT-2000	(TEMBLrel_15,	Created)		
DR	01-OCT-2000	(TEMBLrel_15,	Last sequence update)		
DY	01-MAR-2000	(TEMBLrel_10,	Last annotation update)	/	
DE	Mucin 1 (Fragment).				
DS	Macca maculata (Rhesus macaque).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
CC	Mammalia; Sutheria; Primates; Catarrhini; Cercopithecoidea;				
OX	Cercopithecinae; Macaca.				
RN	NCBI_TaxID=9544;				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=20330533; PubMed=10869775;				
RX	Vaughan H.A., Ho D.W.M., Karanthak V., Sandrin M.S., McKenzie I.F.C.,				
RT	Plesters G.A.,				
RL	"The Immune Response of Mice and Cynomolgus Monkeys to Macaque Mucin1-				
RD	Mannan".				
RE	Vaccine 18:3297-3309(2000) .				
RF	EMBL; AF176947; AAF82403.1;				
RG	InterPro: IPRO01064; CytoStallin_1.				
RI	InterPro: IPRO00082; SEA_domain.				
RJ	Pfam; PFO1390; SEA; 1.				
RS	SMART; SM00200; SEA; 1.				
RU	PROSITE; PSS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.				
RV	PROSITE; PSS0024; SEA; 1.				
RW	NON_TER	1			
RX	FT	1			
RY	NON_TER	553			
SQ	SEQUENCE	553 AA; 55778 MW; 6D7BD6D2EE929318 CRC64;			
Query Match		70.1%; Score 1795.5; DB 6; Length 553;			
Best Local Similarity		67.4%; Pred. No. 9.7e-102;			
Matches 375; Conservative		13; Mismatches 59; Indels 109; Gaps			
21 VTGSGHASTPGEKEXETATONKSVPVSTERNKAVSMNTSVLSHSRGGSSITTGQADPTLT 80					
I VTGSGHNTSPGCEKEIETATORSMISIR-KNAVSMNSR-ISSHVSIVGSSSTTQGQDPVTLL 58					

Qy	81	ADPTSPAGSAAITGQDVTSVVTPEPALGSTTPPAHDVYSA.PDMPK.PPGSTTP.PPHAGTYS	140
Db	59	ALATPEAPAGSATLTGHNVTSAADPTSAAPGSGTGPAGVVTSA.PDTSAA.PGSGTGP.PPVVTS	118
Qy	141	ADPTPEPPGSGTAPPAAGVTSAPDTPAPGSGTAPPAHGVTSAPDMP.PA-----	187
Db	119	ADPTSAAPGSGTGPAPRVTSAPDTPSAADPTSAAPGSGTGPAPVVTSA.PDTSAA.PGSGTGP.PPVVTS	178
Qy	188	-----	187
Db	179	ADPTSAAPGSGTGPAPRVVTSAPDTSAA.PGSGTGPAPRVVTSAPDTSAA.PGSGTGP.PPVVTS	238
Qy	188	-----LGSITP.PPHANTYS.ASGSAGSASTLVH	214
Db	239	ADPTSAAPGSGTGPAPRVVTSAPGTSAA.PGSGTAPPGSGTAPPAHDVTSAPDSAGSASTLVH	298
Qy	215	NGTSRAATTPTPAKSNPPSPBPHSDPTTLTAHSKTKDASTHSHSV.PRLTSSNHSNP	274
Db	299	STTSRAATTPTPAKSNPPSPBPHSDPTTLTAHSKTKDASTHSHSV.PRLTSSNHSNP	357
Qy	275	QJSTGVSEFFSLPRTISNLOPNSLSDPSTDIYQELORLSMF.QIYKCGCLGLSNKF	334
Db	358	QJSLGVSEFFSLPRTISNLOPNSLSDPSTDIYQELORLSMF.QIYKCGCLGLSNMF	417
Qy	335	PRGSVVVOLTAPBEGTINADVFOTFNQYKTEAASRNLTISVSVSHVPPPSAOSGA	394
Db	418	PRGSVVVOSTLVEEGTINADVFOTFNQYKTEAASRNLTISDI.SRVDP.PPSAOTGA	477
Qy	395	GPGGKIALLVLCVLAVALAIYVLLALAVCCQCKRKNYGOLDI.PPARDTYH.PMSEYPTVHT	454
Db	478	GPGGKIALLVLCVLAVALAIYVLLALAVCCQCKRKNYOLDI.PPARDAVYH.PMSEYPTVHT	537
Qy	455	HGRYVPPSSTDRSPEYE 470	
Db	538	HGRYVPPAGGTNRSPYE 553	
RESULT 2			
019115			
ID	019115	PRELIMINARY;	PRT; 602 AA.
AC	019115;		
DT	01-JAN-1998	(TREMblrel. 05, Created)	
DT	01-JAN-1998	(TREMblrel. 05, Last sequence update)	
DE	01-MAR-2002	(TREMblrel. 20, Last annotation update)	
DS	Muc1 (Fragment)		
OS	Oryctolagus cuniculus (Rabbit)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.		
OX	NCBI_TaxId:9986;		
RN	[1]		
RP	SEQUENCE OF 410-602 FROM N.A.		
RC	STRAIN=NEW ZEALAND WHITE.		
RX	MEDLINE=96351712; PubMed=8747930.		
RA	Spicer A.P., Dubig T., Chilton B.S., Gendler S.J.,		
RT	"Analysis of mammalian MUC1 genes reveals potential functionally		
RT	important domains";		
RL	Mamm. Genome 6:885-888(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NEW ZEALAND WHITE.		
RA	Hewelson A., Chilton B.S.;		
RT	"Molecular cloning and hormone-dependent expression of rabbit Muc1 im		
RT	the cervix and uterus";		
RL	Biol. Reprod. 0:0-0(1997).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NEW ZEALAND WHITE.		
RA	Hewelson A., Chilton B.S.;		
RT	Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.		
RT	EMBL; 085787; AAB6380.1;		
RR	Interpict IP8000082; SEA_domain.		
RR	Pfam; PF01390; SEA.1.		

DR SMART; SMO0200; SEA; 1.
 DR PROSITE; PS50024; SEA; 1.
 FT NON TER 1 1
 SO SEQUENCE 602 AA; 61287 MW; 0F4523CF2871E270 CRC64;

Query Match 48.2%; Score 1234; Db 6; Length 602;
 Best Local Similarity 54.7%; Pred. No. 1,7e-67;
 Matches 283; Conservative 42; Mismatches 136; Indels 56; Gaps 10;

QY 22 TSGHGASSTPGGKEKTSATQSSVPS-----STERKVAQMTSSV-----LSH 64
 Db 99 TSTSTSVSDTSPVHEVTSASATSPVHEVTSALTSTSPVSATSPVHEVTSAPATSPVSAT 158
 QY 65 SP-----GGSGTQGGDTLPAPEAPSSAAATGGQVTSVPTTP----- 106
 Db 159 SPVHEVTSAPATSPVSATSPVHEVTSAPATSPVSATSPVHEVTSAPATSPVSAT 218
 QY 107 -----ALGSTTPPAHDVTSAPDNKPAFGSTAPPAHGVTSAPDTPRPPGSGTAPAHGVTS 161
 Db 219 PVHEVTSASATSPVHEVTSAPATSPVSATSPVHEVTSAPATSPVSATSPVHEVTS 278
 QY 162 PSTRPAGSTAPPAHGVTSAPDNKPAFGSTAPPAHGVTSASGSA--SGSASTLVANGTSA 219
 Db 279 LTTSPSTVSATSPVHEVTSAPATSPVSATSPVHEVTSAPATSPVSATSPVHEVTS 338
 QY 220 RATTTP--ASKSTPSPISPEHSDPTPLTASHSTKTDASVTHSHVTPPLTSSNISTPQST 278
 Db 339 LATTPMPVGLSTSTSPSVPS--HTGPPTLTSPS-----HSTGSPTTSTOSTPQVSA 388
 QY 279 GVSFPFSSPFIISLQFNLSLEDSTVDYQGLQDISMPLOIYKQGGFGLGNIFEKPGS 338
 Db 389 GLSPFELSPFTNLQFNLSLEDSSKTYQGLQWRVALLISQYQKPFGLGLGIFRRGS 448
 QY 339 VVVVLTAAPEEGTIVHDEFGQVCKTEASRNLTIDVSVAHPFPSPASGAGVPG 398
 Db 449 VVVDLLAPQESTNKTNYQSGQIQINIQAA--RNINISRTVYADVLPSSAQG--GVPG 505
 QY 399 KQIALALVQVVALATVYLIALANQCRKNGQGLDIPADPTVHMSRYTTHORY 458
 Db 506 KQIALALVQVVALATVYLIALANQCRKNGQGLDIPADPTVHMSRYTTHORY 565
 QY 459 VPPSSTDRSPTEKVSAGKGGSSLTSTNPAVAATSAUL 495
 Db 566 VPPGSTRSPTEKVSAGKGGSGSTYNPAVAATSAUL 602

RESULT 3
 Q8RML4 PRELIMINARY; PRT; 580 AA.
 OS BSMML4;
 AC 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, last annotation update)
 DT MUC1 protein precursor.
 NC MUC1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 BO Bovidae; Bovinae; Bos.
 CX NCBI_TaxID=9913;
 RX [1]
 SN SOURCE FROM N. A.
 RP STRAIN=HOLSTEIN DAIRY COW; TISSUE=LACTATING MAMMARY GLAND;
 RC MEDLINE=21672491; PubMed=1819015
 RA Rasmussen L.K., Andersen M.H., Nielsen R.L., Berglund L.,
 Rasmussen L.K., Petersen T.E., Rasmussen J.J.;
 RT Purification of MUC1 from bovine milk-fat globules and
 RT Characterization of a corresponding full-length cDNA clone.";
 RL J. Dairy Sci. 84:2591-2598(2001).
 DR EMBL; AJ400824; CAC81810.1; -
 DR InterPro; IPR000082; SEA_domain
 DR Pfam; PF01390; SEA_1.
 DR SMART; SMO0200; SEA; 1.

DR PROSITE, PS50024; SEA, 1.
 KW Signal. 1 22 MUC1 PROTEIN.
 FT CHAIN 23 580
 SQ SEQUENCE 580 AA; 58091 MW; E91C13984F7D757 CRC64;
 Query Match 48.0%; Score 1228.5; DB 6; Length 580;
 Best Local Similarity 48.4%; Pred. No. 3.5e-67;
 Matches 285; Conservative 44; Mismatches 157; Indels 103; Gaps 11;

QY 1 MTPGDSPPFLILL---LTVLTVTGSGHASTPGCEKETSATQKRSVSGSTKNAVSM 56
 DB 1 MTPDIOAPFLILLFVLTIVANVPTLTSTSDINPRRTTPTTOSPTSSPTKETSMT 60
 QY 57 TTSVLSSHPGSGSTTQGDVTLAPATEPAGSAAITWQGDVTVPTVRPAL----- 108
 DB 61 TTTLLTASSPASPAPSPGHDGASTPTSPAPSPAPSPGHDGASTPTSPAPSPAPSPG 120
 QY 109 -GSTP-----PAHDVTSAPDNKPAFGSTAPPAHGVTSAPDTRPPGSTA--- 152
 DB 121 DGASTPTSPASPAPSPGHDGASTPTSPAPSPAPSPGHDGASTPTSPAPSPAPSPG 180
 QY 153 -----PAA---HGVTAPDTRPPAGSTAPPAHGVTSAPDNKPAFGSTAPVH 196
 DB 181 DGASTPTSPASPAPSPGHDGASTPTSPAPSPAPSPGHDGASTPTSPAPSPAPSPG 240
 QY 197 NVTSA-----SGSAS 206
 DB 241 NGTSSPTGSPASPPTASPGHDSAPSLTSSPAPSPASPGHDSAPSTSSPTSSMTTSSMS 300
 QY 207 GSASTLVHNGTSARATTTTPASKSTPTSPSHSDPTTLASHTKTDASSTHSTVPT 266
 DB 301 SSWTSAHKGTSSRAWTPTVSKGTPTSVPS--SELTAVTAAHITRTAAS--PSIALST 355
 QY 267 SSNSHTSPDLTGVSPFELSHSNLOFNSSLEDPSTDYVQGLQDSMWLQIYKQSGF 326
 DB 356 SSNPRTSQGLSVSLVSLFSLFRITNLQFNSSLEKPGQSYVQGLQDSMWLQIYKQRF 415
 QY 327 LGLSNKRRPSSVYVCLTAFREGTINHDVEOPNOYKTEAASVYNTLTSQVSNAVF 386
 DB 416 LGLSEIKRRPSSVYVCLTAFREGT-TAEWKAPQSLNMAAS-VNLTISGVSNAVF 473
 QY 387 PPSAOGSAGVGMGIALVYCVYALVYLVLAIVGCCRRKNVGLDIFPARDTYHPM 446
 DB 474 PPSAOGSAGVGMGIALVYCVYALVYLVLAIVGCCRRKNVGLDIFPTLDYHPM 533
 QY 447 SEPTVTHGTVPPSTDSPEYKKSAGNGSGSTSTPAVAATSAVL 495
 DB 534 SEISVTHGTVPPSTDSPEYKKSAGNGSGSTSTPAVAATSAVL 580

RESULT 4
 Q99K60 PRELIMINARY; PRT; 631 AA.
 ID Q99K60;
 AC Q99K60;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Mucin 1, transmembrane.
 GN MUC1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Straube R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC005441; AA05441.1; -;
 DR MGD; MG1:97231; Muc1.
 DR InterPro; IPR000082; SEA-domain.
 DR Pfam; PF01390; SEA; 1.
 DR SMART; SM00200; SEA; 1.

DR PROSITE, PS50024; SEA, 1.
 KW Transmembrane. 631 AA; 64690 MW; B4236DE6655F4C3 CRC64;
 SQ SEQUENCE 631 AA;
 Query Match 46.8%; Score 1198; DB 11; Length 631;
 Best Local Similarity 44.8%; Pred. No. 2.8e-65;
 Matches 284; Conservative 64; Mismatches 144; Indels 142; Gaps 13;

QY 1 MTPGDSPPFLILLITVLT---TVYTGSGHAST---PG 32
 DB 1 MTPGIDAPFLILLIALKGFALPSEBNSVTSQDTSSILASTTPVHSSNDPATRP 60
 QY 33 GKEKTSATQKRSVSGSTKNAVSMTSVLSGSH-----PGSGSTTQGDVTLAPATEP 86
 DB 61 GDSSTPVOSSTSPATRAPEDSTTVALSGTSSPATTPAIVASASPARGDTSRATSL 120
 QY 87 A--SGSAAITWQGDVTVPTVRPALGSTTP-ADHVTAPDNKPAFGSTAPPAHGVTSAPD 143
 DB 121 SKDSNPSPVHSGTSSAPATTAPVDSTSPVHGGTSLSPATSPGDSSTSPHSGTSP 180
 QY 144 TRPPGST-----APAAHGVTSAPDTRPPAGS-TAPPAGVTSA 181
 DB 181 TRAPEDSTTVALSGTSSPATTPAPVDSTSPVHADDTSSPATSPEDSASPVHGGTSS 240
 QY 182 PDNRPALGSTAPVH---VTSAGCSA----- 205
 DB 241 PATSPLDSTSPVHSGASIONIKTSDLASTPDHNGTSVTTSSALGASATSPDHSGTST 300
 QY 206 -----SGSASTLVHNGTSARAT----- 222
 DB 301 TTNSSSDLATTPVYSSMPSTTKVTSGLIIPDHNGSVLPTSSVLGSATSLVNTSAI 360
 QY 223 -TPPAKSTPPSLSHSDPTTLASHTKTDASSTHSTVPTLSSNSHTSPQLSTGV 281
 DB 361 ATPVNGTQSPVSVQVPSPTMAATSSHTSIASSSVYSTVPFSTSSNS-SQLTVGVS 419
 QY 282 PFFLSHINLOFNSSLEDPSTDYVQGLQDSMWLQIYKQSGFGLSNKRRPSSVYV 341
 DB 420 PFFLSHYIQNHFPNSSLSDPSVYQELKRNISGLFLQIF-NGDPLGISIKRRSSVYV 478
 QY 342 QLTAFREGTINHDVEOPNOYKTEAASRYNLISDVSVHPPFPAQSGAGVGMG 401
 DB 479 ESTVFRREGTASVRSQILQHKKE-ADYNTLTISEVKVNMQFPFPAQSRPGVGMG 537
 QY 402 ALVIVCVYALVYLVLAIVGCCRRKNVGLDIFPARDTYHPMSEVPTHTHGRYVP 461
 DB 538 ALVIVCVYALVYLVLAIVGCCRRKNVGLDIFPTQDYHPMSEVPTHTHGRYVP 597
 QY 462 SSTDRSPPEYKKSAGNGSGSTSTPAVAATSAVL 495
 DB 598 GSTDRSPPEYKKSAGNGSGSTSTPAVAATSAVL 631

RESULT 5
 Q95L89 PRELIMINARY; PRT; 554 AA.
 ID Q95L89;
 AC Q95L89;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Mucin (Fragment).
 GN MUC1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OX Bovidae; Bovinae; Bos.
 OX NCBI_TaxId=9913;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Sacchi P.; Rasoeto R.; Rogati S.;
 RL Molecular analysis of MUC1 polymorphism in cattle.
 DT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF399757; AA028023.1; -;

Thu May 8 16:14:05 2003

us-09-658-621b-2.rspt

Page 4

```
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR PROSITE; PSS0024; SEA; 1.
PT NON TER 554
SQ SEQUENCE 554 AA; 55508 MW; E441CD140093AFD CRC64;

Query Match 43.7%; Score 1120.5; DB 6; Length 554;
Best Local Similarity 46.7%; Pred. No. 1.3e-60;
Matches 262; Conservative 41; Mismatches 157; Indels 101; Gaps 10;

QY 1 MTPGSGSFFELLT-----LTVTVTGSGASSTPGCKETSAIORSVPSSTKENVSM 56
DB 1 MTPDQAPFLSTLLLPVLTVANVPLTSDSINFRRTTPVTSTGSPSSPTKETSMST 60
QY 57 TSVVSSHSFGSGSSSTGQGVTLAPTEPASGSAATGQVTVSVTPAL----- 108
DB 61 TTTLLTASAPSPAPSPGHDGASTPTSSPAPSPGHDGASTPTSSPAPSPG 120
QY 109 -GSTP-----PAHDVTAPDNKEAPGSTAPPAHGVTSAPDTPRPPGSTA---- 152
DB 121 DGASTPTSSPAPSPAPSPGHDGASTPTSSPAPSPGHDGASTPTSSPAPSPG 180
QY 153 -----PAA---HGVTAPDTPRAGSTAPPAHGVTSAPDNKPALGSTAPPV 196
DB 181 DGASTPTSSPAPSPAPSPGHDGASTPTSSPAPSPAPSPGHDGASTPTSSPAPSPG 240
QY 197 NVTSA-----SSSAS 206
DB 241 NGTSPPTGSPAPSPAPSPGHDGASTPTSSPAPSPAPSPGHDGASTPTSSPTSMVTGMS 300
QY 207 GSATLVHNTGTSARATTPPAKSTPPSPISGHSUTPTLASHTKTPDASTHSTVPLT 266
DB 301 SSMVTSAHKTSRAATTPVSKGTPSSVPS--SRAAPTAASHITRTPAS--PSIALST 355
QY 267 SSNSTSPOLSTGVSPFSPHSINLOPNSLSDPSTGYOELORISMPLOYKGGC 326
DB 356 SSNPSTQGLSVSVLYFSFRINLOPNSLSDPSTGYOELORISMPLOYKGGC 415
QY 327 LGLNIFKRGSVVQTLAFLRGTINVDVETONOKYKTEASRNLTISVSVSHVP 386
DB 416 LGLSEIFKRGSVVVELTARREGT-PAEWYKAGQSLEAHMS-YNLTLSGVSVTSAFP 473
QY 387 PPSNOSGAGVPGKIALVLCVVALATVYLAALVAGCCRRKNGQLDIPADPTHEM 446
DB 474 PPSNOSGAGVPGKIALVLCVVALATVYLAALVAGCCRRKNGQLDIPADPTHEM 533
QY 447 SEVPTHTGRVYPPGSTDRS 467
DB 534 SEVPTHTGRVYPPGSTDRS 554

RESULT 6
ID 060551 PRELIMINARY; PRT; 193 AA.
AC 060551.
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE Epithelial mucin (Fragment).
GN MUC1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE=96351712; PubMed=8747930;
RA Splicer A.P.; Duhig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains.";
RL Mamm. Genome 6:885-888(1995).
```

```
DR EMBL; L41545; AAB48541.1; -.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PSS0024; SEA; 1.
PT NON TER 1
SQ SEQUENCE 193 AA; 21201 MW; 635CF01D27BC46BA CRC64;

Query Match 28.1%; Score 720.5; DB 11; Length 193;
Best Local Similarity 70.9%; Pred. No. 8.5e-37;
Matches 139; Conservative 25; Mismatches 29; Indels 3; Gaps 3;

QY 300 DPSDYYOELORISMEFLQYKGGFLGSLNIFKRGSVVQTLAFLRGTINVDVET 359
DB 1 DSSNYYOELKRVSGSLFLQYFSR-APLQISTIFKRGSVVQDVTIFRSGAVNASEVKS 59
QY 360 QNOKYTEASRYNLTISDVSVHPPPSAGSGVPGKIALVLCVVALATVYLA 419
DB 60 QLTGHEGE-AEYNALISKIVGEMQPPSSAGSMGVPNGIALVLCVVALATVYLA 118
QY 420 ALAVCCRRKNGQLDIPADPTYHMSSEPTHTGRVYPPGSTDRSPYKVSAGNGCS 479
DB 119 ALAVCCRRKNGQLDIPADPTSHMSSEPTHTHTGRVYPPGSTDRSPYKVSAGN-GS 177
QY 480 SLSTYNAVAATSANL 495
DB 178 SLSTYNEVAATISANL 193

RESULT 7
ID 028078 PRELIMINARY; PRT; 192 AA.
AC 028078.
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE Epithelial mucin (Fragment).
GN MUC1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE=LACTATING MAMMARY GLAND;
RX MEDLINE=96351712; PubMed=8747930;
RA Splicer A.P.; Duhig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains.";
RL Mamm. Genome 6:885-888(1995).
DR EMBL; L41543; AAB48538.1; -.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PSS0024; SEA; 1.
PT NON TER 1
SQ SEQUENCE 192 AA; 20930 MW; 01DFFA618079715E CRC64;

Query Match 27.9%; Score 716; DB 6; Length 192;
Best Local Similarity 74.5%; Pred. No. 1.6e-36;
Matches 146; Conservative 12; Mismatches 34; Indels 4; Gaps 3;

QY 300 DPSDYYOELORISMEFLQYKGGFLGSLNIFKRGSVVQTLAFLRGTINVDVET 359
DB 1 NQTSYYOELORSINGLTLQYKQRFGLSEIKFRPGSVVVELTARREGT-PAEWYNA 59
QY 360 QNOKYTEASRYNLTISDVSVHPPPSAGSGVPGKIALVLCVVALATVYLA 419
DB 60 QPSOLEHMAAS-YNLTISGVSVSAPFPSSAQAGVPGKIALVLCVVALATVYLA 118
QY 420 ALAVCCRRKNGQLDIPADPTYHMSSEPTHTGRVYPPGSTDRSPYKVSAGNGCS 479
DB 420 ALAVCCRRKNGQLDIPADPTYHMSSEPTHTGRVYPPGSTDRSPYKVSAGNGCS 479
```

```

Db 119 ALVCCCRKKGGLDIFPRLDHYHPSSEVYTHHGRVVPSTKSPYEVSAGNGS 178
Qy 480 SLSTYNPAVAATSAHL 495
Db 179 SLSTYNPAVAATSAHL 192

RESULT 8
Q28723 PRELIMINARY, PRT, 193 AA.
ID Q28723
AC Q28723,
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Epithelial mucin (Fragment).
GN MUC1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE=96351712; PubMed=8747930;
RA Spicer A.P., Duhig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains."
RL Mamm. Genome 6:885-888(1995).
DR EMBL; L41544; AAB48540.1; -.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER
SQ SEQUENCE 193 AA; 21069 MW; POC31DA4D98C80B9 CRC64;

Query Match 27.6%; Score 706.5; DB 6; Length 193;
Best Local Similarity 73.3%; Pred. No. 6, 1e-36;
Matches 143; Conservative 15; Mismatches 34; Indels 3; Gaps 2;

Qy 301 PESTDYQELQDISEMFLQIYKQGFGLSNIKRPGSVVQLTAFBGTINVDHVT 360
Db 2 PESTKYQELQNVASALISQIYKQKFLSGIRKRSQSVGVLLAQSSTHVTYVRSQ 61
Qy 361 FNYKTEASRYNLTIDSVSHVPPPSAQSAGVPGMGIALLVLCVLAIVYL 420
Db 62 LQNIHQAA-RNINLISRYTVRDVLSFSSAQS--GVPGMGIALLVLCVLAIVYL 118
Qy 421 LAVCCCRKKGGLDIFPARDTYHPSSEVYTHHGRVVPSTKSPYEVSAGNGSS 480
Db 119 LAVCCCRKKGGLDIFPARDTYHPSSEVYTHHGRVVPSTKSPYEVSAGNGSSG 178
Qy 481 SLSTYNPAVAATSAHL 495
Db 179 SLSTYNPAVAATSAHL 192

RESULT 9
Q60408 PRELIMINARY, PRT, 193 AA.
ID Q60408
AC Q60408;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Epithelial mucin (Fragment).
GN MUC1.
OS Cavia cutleri (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognath; Caviidae; Cavia.
OX NCBI_TaxID=10144;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LACTATING MAMMARY GLAND;

```

```

RX MEDLINE=96351712; PubMed=8747930;
RA Spicer A.P., Duhig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains."
RL Mamm. Genome 6:885-888(1995).
DR EMBL; L41544; AAB48540.1; -.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER
SQ SEQUENCE 193 AA; 21661 MW; D2593E4EB9ECL2F CRC64;

Query Match 27.3%; Score 699.5; DB 11; Length 193;
Best Local Similarity 70.9%; Pred. No. 1, 6e-35;
Matches 139; Conservative 20; Mismatches 34; Indels 3; Gaps 3;

Qy 300 PESTDYQELQDISEMFLQIYKQGFGLSNIKRPGSVVQLTAFBGTINVDHVT 359
Db 1 PESTRYQELQKNITRLFLQIYQO-DFLGTLNIRKPGSVAVESTVIFRNNAVESFVS 59
Qy 360 QFNQYTEASRYNLTIDSVSHVPPPSAQSAGVPGMGIALLVLCVLAIVYL 419
Db 60 QLTQK-EDATRYMLVISEVAREVQVT-SSTAQSGVPGMGIALLVLCVLAIVYL 117
Qy 420 ALVCCCRKKGGLDIFPARDTYHPSSEVYTHHGRVVPSTKSPYEVSAGNGSS 479
Db 118 ALVCCCRKKGGLDIFPARDTYHPSSEVYTHHGRVVPSTKSPYEVSAGNGS 177
Qy 480 SLSTYNPAVAATSAHL 495
Db 178 SLSTYNPAVAATSAHL 193

RESULT 10
Q35770 PRELIMINARY, PRT, 74 AA.
ID Q35770
AC Q35770;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Mucin 1 (Fragment).
GN MUC1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96285070; PubMed=9623612;
RA Mesquita M.M., Mahi S.K., Julian J., Carson D.D.;
RT "Reduction of mucin-1 expression during the receptive phase in the rat
RT uterus."
RL Biol. Reprod. 58:1503-1507(1998).
DR EMBL; AF007544; AAB82948.1; -.
FT NON_TER
SQ SEQUENCE 74 AA; 8187 MW; 12255FB4740D9EAF CRC64;

Query Match 14.0%; Score 358; DB 11; Length 74;
Best Local Similarity 85.1%; Pred. No. 3, 4e-15;
Matches 63; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 422 AVCCCRKKGGLDIFPARDTYHPSSEVYTHHGRVVPSTKSPYEVSAGNGSSL 481
Db 1 AVCCCRKKGGLDIFPARDTYHPSSEVYTHHGRVVPSTKSPYEVSAGNGSSL 60
Qy 482 SYTNPAVAATSAHL 495
Db 61 SYTNPAVAATSAHL 74

RESULT 11
Q92B39

```

```

ID Q92B39 PRELIMINARY; PRT; 487 AA.
AC Q92B39;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Multiple banded antigen.
OS Ureaplasma urealyticum (Ureaplasma urealyticum biotype 2).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
NCBI_TaxID=2130;
RN [1]
RP SEQUENCE FROM N.A.
RA Zheng X., Teng L.-J., Watson H.L., Glass J.I., Cassell G.H.;
RT "Sequence analysis of the major serovar specific antigen (M8 antigen)
RT genes from different serovars of Ureaplasma urealyticum,"
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50459; AAD0075.1; -.
DR InterPro; IPR003302; Cornifin.
DR Pfam; PF02389; Cornifin; 4.
SQ SEQUENCE 487 AA; 46665 MW; B901856DADB3724 CRC64;

Query Match 10.4%; Score 267.5; DB 2; Length 487;
Best Local Similarity 32.5%; Pred. No. 97e-09;
Matches 100; Conservative 28; Mismatches 117; Indels 63; Gaps 16;

QY 3 PCTGSPFLLLLLVTVVSSGASSTGCKET-----SATGRSSVSSSTERNVMSMT 57
DB 149 PGGSS-----TTPGSG-STTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSG 195
QY 58 SSVLSHSPGSGSTTGQDVTLP-----ATPAGSANTGQDVTSP--VTPPALG 109
DB 196 -----QPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSG 248
QY 110 STTPPAHVTSAPDN-----XPAGSTAPPAHVTSAP-----DTPPGSTAPAAVTS 161
DB 249 STTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSG 308
QY 162 P-----DTPPAHVTSAPDN-----RPAIGSTAPPAHVTSASGAS-----GSA 209
DB 309 PGGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSG 368
QY 210 STLVNHTSARATTPPAKSTPSPSPHSHDPTTTLASHSTKTDASTN--HSTVPLLT 266
DB 369 ST-----TTPGSGSTTPGSGST--TTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSG 421
QY 267 SSNHTSP 274
DB 422 GSGSTTP 429

RESULT 12
ID Q9RKR9 PRELIMINARY; PRT; 1334 AA.
AC Q9RKR9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative multi-domain regulatory protein.
GN SC02259 OR SC075A.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parthill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Krensch H., Kleser H.M., Denaplatte D., Richner A., Cullum J.,
RA Rabinowitsch B., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb streptomycetes coelicolor A3(2) chromosome,"
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.F., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Layke I., Murphy J., Oliver K., O'Neill S.,
RA Rabinowitsch B., Rajandream M.A., Rutherford K., Butler S., Taylor S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parthill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2),"
RL Nature 411:141-147(2002).
DR EMBL; AL133201; CAB61705.1; -.
DR InterPro; IPR005158; BAD.
DR InterPro; IPR00767; Disease_resist.
DR InterPro; IPR001867; Trans_reg_C.
DR Pfam; PF03704; BAD; 1.
DR PRINTS; PR00364; DISBASERS1ST.
DR PRINTS; PR01574; TUBSFPROTEIN.
DR Prodom; PD000329; Trans_reg_C; 1.
SQ SEQUENCE 1334 AA; 138787 MW; 78DC74683E8778C CRC64;

Query Match 10.4%; Score 265.5; DB 16; Length 1334;
Best Local Similarity 33.7%; Pred. No. 4.1e-08;
Matches 91; Conservative 14; Mismatches 90; Indels 75; Gaps 12;

QY 31 PGGKETSATGRSSVSSSTERNVMSMTSSVLSHSPGSGSTTGQDVTLPATPEPAGS 90
DB 237 PGPRLRTMAELLS-PSPTPTGSRSTPTGWTGPGAGAGAAAGTGV-----ASGA 287
QY 91 AATWGQDVTSPVTRPAL-----GSTTPPAHVTSAPDNKPAAG- 129
DB 288 GAAGGPPAGSPAGSAPAAVAGSGGCGAPGCMWPAPCTAGPAGSTAPPHDTAAADAPAPGP 347
QY 130 -----STAPPAHVTSAPDTR--PPPGS-----TAPPAHVTSAPDTPAPAGSTAP- 173
DB 348 TSAPGTAPAAAGTAPAAAGTAPAGTAPAGTAPAAVAGTAPAAVAGTAPAGTAPAGTAPAGTAP 406
QY 174 -----PAHVTSAPDNRPALGSTAPPAHVTSASGSA-----SGSASTLVH 214
DB 407 AAGTAPAAAGTAPAAAGTAPAGTAPAGTAPAGTAPAAVAGTAPAAVAGTAPAGTAPAGTAPAG 460
QY 215 NCTSARATTPP--ASKSTPSPSPHSHDTP 242
DB 461 -CTGCAATPPFAAAAGAGSAPAPAPGP 489

RESULT 13
ID Q9N4S7 PRELIMINARY; PRT; 1079 AA.
AC Q9N4S7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Y51B11A.1 protein.
GN Y51B11A.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Rhabditinae; Caenorhabditis.
NCBI_TaxID=6235;
RN [1]

```

RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE93069613; PubMed9851916;
 RA None;
 RT "Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium."
 RL Science 282:2012-2018 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX Cotton M.;
 RA "The sequence of *C. elegans* cosmid Y51B1A."
 RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX Waterston R.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL, AC006797; AAF60743.1; .
 DR InterPro; IPR002965; P. Fitch; extensn.
 DR PRINTS; PR01217; PRICHEXTENS.
 SQ SEQUENCE 1079 AA; 110532 MW; 8BDB3824CF80CA1 CRC64;
 Query Match 10.3%; Score 264.5; DB 5; Length 1079;
 Best Local Similarity 32.0%; Pred. No. 3,7e-08;
 Matches 96; Conservative 35; Mismatches 126; Indels 41; Gaps 13;
 QY 19 TVVTSGHSSTPGGKETSATQSSV--PSS--TEKNVSVMTSSVLSHSPGSGSTTQ 74
 DB 520 TSTSPSSSTTPVQTTTAPETTTSTPSSSTTPVQTTTAPETTTSTPSSSTTPV 579
 QY 75 GQDVLAP--ATEPAGSAAATGQDVTSV--VTRPALGSTTPADVTASAD--NK 125
 DB 580 QTTTATETTTSTPSSSTTPVQTTTAPETTTSTPSSSTTPVQTTTAPETTTST 639
 QY 126 PARGSTAPPAAGTASAP--TRPPGSTAPAGTASAP--TRPAGSTAPAGT 179
 DB 640 PPSSTTPVQTTTAPETTTSTPSSSTTPVQTTTAPETTTSTPSSSTTPVQTTT 699
 QY 180 SAPD--NRPALGSTAPVHNTVTSAGSAS--GSASTLVHNGTSARATTPASKST 231
 DB 700 TAPETTTSTPSSSTTPVQTTTAPETTTSTPSSSTTPVQTTTAPETTTSTPSS 755
 QY 232 PSLPHSHDTPTLASHSTKTDASS-----THASTVPLTSS--NHSTPOLSTGV 281
 DB 756 PPSSTTPVQTTTAPETTTSTPSSSTTPVQTTTAPETTTSTPSSSTTPVQTTT 815
 RESULT 14
 ID 08S037 PRELIMINARY; PRT; 160 AA.
 AC 08S037;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DS Nuclein (fragment).
 OS Bos taurus (bovine).
 GN MUC1/VNTR2.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HOLSTEIN;
 RX Yamamoto N.;
 RA "Bos taurus MUC1 VNTR (Type2)."
 RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AB084274; BAB9140.1; .
 FT NON_TER 1 160
 SQ SEQUENCE 160 AA; 14178 MW; 61CG8305BC403B03 CRC64;

Query Match 9.9%; Score 254.5; DB 6; Length 160;
 Best Local Similarity 36.8%; Pred. No. 1,7e-08;
 Matches 60; Conservative 14; Mismatches 82; Indels 7; Gaps 3;
 QY 71 STTGQDVTTLAPATEPAGSAAATGQDVTSVPTPALGSTTPPAHDVTSAAPDKTAP 130
 DB 3 AASPGHDASTPTSSPAPSPAPSPGHDASTPTSSPAPSPAPSPGHDASTPTSSPAP 62
 QY 131 TAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDNR 190
 DB 63 AASPGHDASTPTSSPAPSPAPSPGHDASTPTSSPAPSPAPSPGHDASTPTSSPAP 122
 QY 191 TAPPVHNTVSA--SGS--ASGASTLVHNGTSARATTPASKSTP 231
 DB 123 AASPGHGTSSPTGSPAPSPAPSPGHDAS----PTSSPAP 160
 RESULT 15
 ID 09BK77 PRELIMINARY; PRT; 1325 AA.
 AC 09BK77;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DS PPG3.
 GN PPG3.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OC NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FRIEDLIN;
 RA Wyler P.J., Sisk E., Cawthra J., Handley F., Vogt C., Robertson L., McDonagh P., Ivens A., Nguyen D., Munden H., Stuart K., Worthey E.A.;
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AC094329; AAK31375.1; .
 DR InterPro; IPR001611; LRR_TYP.
 DR InterPro; IPR003591; LRR_TYP.
 DR Pfam; PF00360; LRR_8; LRR_TYP.
 DR PRINTS; PR00019; LEURICHRPT.
 DR SMART; SM00369; LRR_TYP; 5.
 SQ SEQUENCE 1325 AA; 133855 MW; 8898928BAE9418F CRC64;
 Query Match 9.9%; Score 253.5; DB 5; Length 1325;
 Best Local Similarity 27.2%; Pred. No. 2,2e-07;
 Matches 97; Conservative 65; Mismatches 143; Indels 51; Gaps 12;
 QY 27 AASTPGGKETSATORSSVSTTEKNVSVMTSSV--LSHSPGSGSTTQDVTTLAPAT 84
 DB 834 SSGAPASASSAPSSSSAPASASSAPSSSSAPASASSAPSSSSAPASASSAPSS 893
 QY 85 EPASGSAATMGQDVTSVPTPALGSTTPPAHDVTSAAP--DNKPARGSTAPPAHGV 140
 DB 894 SSSSAPASASSAPSSSSAPASASSAPSSSSAPASASSAPSSSSAPASASSAPSS 953
 QY 141 APTTRPPGSTAPPAHGVTSAP--DTRPAGSTAPPAHGVTSAP--DNKRALGSTAP 193
 DB 954 APSS--SSAPASASS--SSAPASASS--SSAPASASS--SSAPASASS--SSAP 1008
 QY 194 PVHNTVTSAGSAGSASTLVHNGTSARATTPASKSTP--GTPSHSDTPVTTLAS 247
 DB 1009 -----SASSSAPSSSS--SAPASASSAPSSSSAPASASSAPSSSSAPAS 1059
 QY 248 --HSTVDSASTHSTVPTPLTSSNHSTSPOLSTGVSTFFLSPIHNLQFNLSLDPET 305
 DB 1060 APSSSSSAPASASSAPSSSSAPASASSAPSSSSAPASASSAPSSSSAPAS 1114
 QY 306 YQELQDUSSEMFQ-----IYKGGFLGLSVIKPRQSVVQUTLAPFPG 350
 DB 1115 YPTQVAHSMFLGFSFESFADASATWKCIPFCMAGITCRPBGFSFLGAVVMSG 1170
 Search completed: May 1, 2003, 23:12:47

Thu May 8 16:14:05 2003

us-09-658-621b-2.rpt

Page 8

Job time : 110.089 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 22:59:30 ; Search time 17.6786 Seconds
(without alignments)
1161.337 Million cell updates/sec

Title: US-09-658-621B-2

Sequence: 1 MPTGTQSPFFLLTLTLTV.....NGSSLSYTNPAVAANTSANL 495

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2234	87.2	475	1 MUC1_HYLLA	Q29435 hylobates 1
2	2130	83.1	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
3	1207.5	46.9	630	1 MUC1_MOUSE	O02496 mus musculus
4	1156.5	45.1	676	1 MUC1_MESAU	O60528 mesocricetu
5	236.5	9.2	907	1 VGP3_BBV	P03200 epstein-bar
6	234	9.1	1162	1 TON4_TRYCR	P23253 trypanosoma
7	233.5	9.1	1367	1 AMY1_YEAST	P06640 saccharomyc
8	233	9.1	3178	1 VSG9_CAEU	O09284 caenorhabdi
9	229	8.9	886	1 VGP3_BBV28	O07284 epstein-bar
10	225	8.8	725	1 AGA1_YEAST	P32323 saccharomyc
11	221	8.6	797	1 VGL1_HYERB	P28968 equine hepr
12	216.5	8.5	5179	1 MUC2_HUMAN	O02817 homo sapien
13	209.5	8.1	1161	1 DAN4_YEAST	P47179 saccharomyc
14	208.5	8.1	573	1 CL14_MOUSE	P19467 mus musculus
15	206.5	8.1	605	1 YHAC_YEAST	P38739 saccharomyc
16	205	8.1	598	1 YHAF_SCHRO	O10168 schizosacch
17	200	7.8	610	1 MUC4_HUMAN	O09102 homo sapien
18	189.5	7.4	551	1 BDD4_BABIT	O28645 oryzocollig
19	186	7.3	1152	1 MAP4_HUMAN	P27813 homo sapien
20	183.5	7.2	1191	1 N121_XET	P52581 xeroderma
21	182	7.1	400	1 MUC1_XETLA	P14583 xeroderma
22	180	7.0	634	1 VEGF1_CANAL	P13533 canalis alb
23	180	7.0	634	1 VEGF2_CANAL	P13533 canalis alb
24	176	7.0	721	1 YH82_MYCTU	O10330 mycobacteri
25	176	6.9	670	1 VSS0_HSV1	O00430 iccatalurid
26	176	6.9	1083	1 T2D3_HUMAN	O00430 iccatalurid
27	176	6.9	1229	1 N121_HUMAN	O00430 iccatalurid
28	175	6.8	530	1 YH9A_SCHRO	O09188 schizosacch
29	174	6.8	283	1 EYTN_SCHRO	O09188 schizosacch
30	174	6.8	1115	1 NCAL_MOUSE	P13355 mus musculu
31	174	6.8	1125	1 MAP4_MOUSE	P27546 mus musculu
32	174	6.8	1140	1 YH96_YEAST	O04893 saccharomyc
33	174	6.8	1185	1 DRP6_HUMAN	P54253 homo sapien

34	173.5	6.8	1251	1 YG03_CAEU	O09550 caenorhabdi
35	172.5	6.7	1072	1 MAP4_BOVIN	P16225 bos taurus
36	171	6.7	636	1 YH9A_YEAST	P13882 saccharomyc
37	170.5	6.7	528	1 FOD4_HUMAN	O00597 homo sapien
38	169.5	6.6	526	1 K062_MOUSE	O63856 mus musculu
39	169	6.6	1150	1 APMD_PIG	P13021 sus scrofa
40	168	6.6	817	1 VRF1_YEAST	P33750 saccharomyc
41	168	6.6	1589	1 PHP_DROME	P33765 drosophila
42	167.5	6.5	378	1 LSKX_RAT	P13328 rattus norv
43	167	6.5	600	1 SP96_DICDI	P13328 rattus norv
44	166	6.5	567	1 CH13_CANAL	P40954 canalis alb
45	166	6.5	786	1 STJ1_DROME	O05319 drosophila

ALIGNMENTS

RESULT 1	ID	MUC1_HYLLA	STANDARD:	PRT:	475 AA.
AC	Q29435;	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DE	15-JUN-2002 (Rel. 41, Last annotation update)				
GN	Mucin 1 precursor (MUC-1).				
OS	Hylobates lar (Common gibbon).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.				
OX	NCBI_TaxID=9580;				
RY	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=96351712; PubMed=8747930;				
RA	Spicer A.P., Dubig T., Chilton B.S., Gendler S.J.;				
RT	"Analysis of mammalian MUC1 genes reveals potential functionally				
RT	important domains";				
RL	Mamm. Genome 6:885-888(1995).				
CC	- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN				
CC	- CYTOSKELETON (BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: Type 1 membrane protein.				
CC	- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.				
CC	- SIMILARITY: CONTAINS 1 SEA DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: L41589; AAA69965.1; -				
DR	EMBL: L41625; AAA69918.1; JOINED.				
DR	EMBL: L41624; AAA69918.1; JOINED.				
DR	InterPro: IPR000082; SEA_domain.				
DR	Pfam: PF01390; SEA; 1.				
DR	SMART: SM00200; SEA; 1.				
DR	PROSITE: PS50024; SEA; 1.				
KW	Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;				
KW	Repeat.				
FT	SIGNAL	1	23	POTENTIAL.	
FT	CHAIN	24	475	MUCIN 1.	
FT	DOMAIN	24	380	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM	381	401	POTENTIAL.	
FT	DOMAIN	402	475	CYTOPLASMIC (POTENTIAL).	
FT	REPEAT	102	121	1.	
FT	REPEAT	122	141	2.	
FT	REPEAT	142	161	3.	
FT	REPEAT	162	181	4.	
FT	DOMAIN	254	371	SEA.	
SQ	SEQUENCE	475 AA;	49371 MW;	D7A699D6D8C6622 CRC64;	

Query Match 87.2%; Score 2234; DB 1; Length 475;

Best Local Similarity 99.1%; Fred. No. 1.3e-100;
Matches 441; Conservative 8; Mismatches 26; Indels 20; Gaps 1;

QY 1 MTPGTOGPPFLLLLTTCVTTTGTSGHASTPGGKEKETSATQSSVSPSTTEKNVMTSSV 60
Db 1 MTPGQGLPFLLLLTTCVTTTGTSGHASTPGGKEKETSATQSSVSPSTTEKNVMTSSV 60
QY 61 LSSHPGSGSTTGGQDVTALPATERPSSAAWQGVTVVPTRPALGCTTPPADVTS 120
Db 61 LSSHPGSGSTTGGQDVTALPATERPSSAAWQGVTVVPTRPALGCTTPPADVTS 120
QY 121 APDNKPAVPGSTAPPAHVTSAPDTRPPGCTAPAAVGTSAEDTPRPAAGTAPPAHVTS 180
Db 121 APDNKPAVPGSTAPPAHVTSAPDTRPPGCTAPAAVGTSAEDTPRPAAGTAPPAHVTS 180
QY 181 APDNKPAVPGSTAPPAHVTSAPDTRPPGCTAPAAVGTSAEDTPRPAAGTAPPAHVTS 240
Db 181 APDNKPAVPGSTAPPAHVTSAPDTRPPGCTAPAAVGTSAEDTPRPAAGTAPPAHVTS 240
QY 241 TPTTLASHSTKTDAASSTHSTVPPPLTSSNHSSTPQSTGVSPFPLSFIHSMLOFNSSLED 300
Db 241 TPTTLASHSTKTDAASSTHSTVPPPLTSSNHSSTPQSTGVSPFPLSFIHSMLOFNSSLED 300
QY 301 PSTDYVQELQSDISEMFLQIYKQGFGLSNIKFPGSVVVOULTAFREGTTNVHDEVTO 360
Db 301 PSTDYVQELQSDISEMFLQIYKQGFGLSNIKFPGSVVVOULTAFREGTTNVHDEVTO 360
QY 361 PNQYKTAASRNLTISDVSVSHVPPFPGSQAGVPGMGIALVYCVLVALATVYLLA 420
Db 361 PNQYKTAASRNLTISDVSVSHVPPFPGSQAGVPGMGIALVYCVLVALATVYLLA 420
QY 421 LAVQCCRRKRYGQLDIPARPTVPMSEVPTVTHGRVPSSTDRSPYKVSAGKSSS 480
Db 421 LAVQCCRRKRYGQLDIPARPTVPMSEVPTVTHGRVPSSTDRSPYKVSAGKSSS 480
QY 481 LSYNPAVAATSAANL 495
Db 481 LSYNPAVAATSAANL 495
QY 495 LSYNPAVAATSAANL 495
Db 495 LSYNPAVAATSAANL 495

RESULT 2
MUC1_HUMAN STANDARD; PRT; 1255 AA.
AC P15941; P15942; P13931; P17626; Q14128; Q16437; Q9Y4J2;
AC Q16615; Q14876; Q9UE75; Q9UE76; Q9UOL1; Q9BXA4;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)
DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)
DE (Tumor-associated epithelial membrane antigen) (EMA) (H23AG) (Peanut-
DE reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen
DE DF3) (CD227 antigen).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eularchia; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RN SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Pancreas;
RA MEDLINE=90368716; PubMed=2394722;
RA Ian M.S., Bata S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA";
RT J. Biol. Chem. 265:15294-15299 (1990).
RN 12
RN SEQUENCE FROM N.A. (ISOFORMS A AND B).
RA MEDLINE=90202794; PubMed=2318825;
RA Lidgenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilken J.;
RT "Episialin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini";
RT J. Biol. Chem. 265:5573-5578 (1990).

RN 13
RN SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Breast carcinoma;
RA MEDLINE=90368715; PubMed=1697589;
RA Gendler S.D., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Peet N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin";
RT J. Biol. Chem. 265:15286-15293 (1990).
RN 14
RN SEQUENCE FROM N.A. (ISOFORM A).
RA MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Peet N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.D.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit";
RT Biochem. Biophys. Res. Commun. 173:1019-1029 (1990).
RN 15
RN SEQUENCE FROM N.A. (ISOFORM SRC).
RC TISSUE=Breast carcinoma;
RA MEDLINE=90276413; PubMed=2351132;
RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J.,
RA Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms";
RT Eur. J. Biochem. 189:463-473 (1990).
RN 16
RN SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RA MEDLINE=90276414; PubMed=2112460;
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA Zaretsky J., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transcribed gene and over-expression in breast cancer
RT tissue";
RT Eur. J. Biochem. 189:475-486 (1990).
RN 17
RN SEQUENCE FROM N.A. (ISOFORM A).
RA MEDLINE=91031045; PubMed=1688329;
RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
RA Jeltsch J.W., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen";
RT Gene 93:313-318 (1990).
RN 18
RN SEQUENCE FROM N.A. (ISOFORM Y).
RA MEDLINE=95010060; PubMed=7925397;
RA Zrihan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,
RA Keydar I., Hilken J., Wreschner D.H.;
RT "Characterization and molecular cloning of a novel MUC1 protein,
RT devoid of tandem repeats, expressed in human breast cancer tissue";
RT Eur. J. Biochem. 224:787-795 (1994).
RN 19
RN SEQUENCE FROM N.A. (ISOFORMS X; Y AND Z).
RA MEDLINE=97355747; PubMed=9212288;
RA Oosterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,
RA Finstad C.L.;
RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
RT cancer cell lines and demonstration of a new short variant form
RT (MUC-1/Z)";
RT Int. J. Cancer 72:87-94 (1997).
RN 110
RN SEQUENCE FROM N.A. (ISOFORM Y).
RA Zhang L.X., Li C.H.;
RT "Molecular cloning of an isoform of MUC1, MUC1/Y";
RT Submitted (FEB-1999) to the EMBL/GenBank/DBD databases.
RN 111
RN SEQUENCE FROM N.A. (ISOFORM S).
RA Zhang L.X., Li C.H., Sun L.Y., Yue W.;
RT "Cloning of a new potential secreted short variant form of MUC1 mucin
RT in epithelial cancer cell line";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [12]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=88330762; PubMed=3417635;
 RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
 RA Burchell J.,
 RT "A highly immunogenic region of a human polymorphic epithelial mucin
 RT expressed by carcinomas is made up of tandem repeats."
 RL J. Biol. Chem. 263:12820-12823(1988).
 RN [13]
 RP SEQUENCE OF 1-160 FROM N.A. (ISOFORM B).
 RX MEDLINE=9008473; PubMed=2597151;
 RA Abe M., Siddiqui J., Kufe D.,
 RT "Sequence analysis of the 5' region of the human DF3 breast
 RT carcinoma-associated antigen gene."
 RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
 RN [14]
 RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM B).
 RX TISSUE=Thyroid;
 MEDLINE=96181746; PubMed=8608966;
 RA Weiss M., Baruch A., Keydar I., Wreschner D.H.,
 RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
 RT transcriptase polymerase chain reaction of the MUC1 gene."
 RL Int. J. Cancer 66:55-59(1996).
 RN [15]
 RP SEQUENCE OF 1-89 FROM N.A.
 RX TISSUE=Lung;
 MEDLINE=96181746; PubMed=8604237;
 RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
 RA Lee L.N., Luh K.T., Wu C.W.,
 RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
 RT tissues."
 RL Oncology 53:118-126(1996).
 RN [16]
 RP SEQUENCE OF 1-46 FROM N.A. (ISOFORMS C AND D).
 RX TISSUE=Breast carcinoma;
 Bulwella L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.,
 RA Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
 RX MEDLINE=97460054; PubMed=9312074;
 RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,
 RA Hantsch F.-G.,
 RT "Localization of O-glycosylation sites on glycopeptide fragments from
 RT lactation-associated MUC1. All putative sites within the tandem
 RT repeat are glycosylation targets in vivo."
 RL J. Biol. Chem. 272:24780-24793(1997).
 RN [18]
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
 RX MEDLINE=99303572; PubMed=10372415;
 RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,
 RA Hantsch F.-G.,
 RT "High density O-glycosylation on tandem repeat peptide from secretory
 RT MUC1 of T47D breast cancer cells."
 RL J. Biol. Chem. 274:18165-18172(1999).
 RN [19]
 RP POLYMORPHISM WITHIN THE REPEAT.
 RX MEDLINE=21359366; PubMed=11350974;
 RA Engelmann K., Belaud S.R., Hantsch F.-G.,
 RT "Identification and topology of variant sequences within individual
 RT repeat domains of the human epithelial tumor mucin MUC1."
 RL J. Biol. Chem. 276:27764-27769(2001).
 RN [20]
 RP CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
 RX MEDLINE=99211485; PubMed=10197628;
 RA Baruch A., Hartmann M.-L., Yoeli M., Adereth Y., Greenstein S.,
 RA Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,
 RA Wreschner D.H.,
 RT "The breast cancer-associated MUC1 gene generates both a receptor and
 RT its cognate binding protein."
 RL Cancer Res. 59:1552-1561(1999).
 RN [21]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.

RX MEDLINE=21240104; PubMed=11341784;
 RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
 RA Harris A.,
 RT "Identification of MUC1 proteolytic cleavage sites in vivo."
 RL Biochem. Biophys. Res. Commun. 283:715-720(2001).
 RN [22]
 RP CHARACTERIZATION.
 RX MEDLINE=21836452; PubMed=11847293;
 RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,
 RA Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,
 RA Stacey M., Lin H.-H., Gordon S.,
 RT "Generation of ligand-receptor alliances by 'SPA' module-mediated
 RT cleavage of membrane-associated mucin proteins."
 RL Protein Sci. 11:698-706(2002).
 RN [23]
 RP PHOSPHORYLATION.
 RX MEDLINE=95080414; PubMed=7988707;
 RA Zrihan-Licht S., Baruch A., Elroy-Stein O., Keydar I., Wreschner D.H.,
 RT "Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins.
 RT Cytokine receptor-like molecules."
 RL FEBS Lett. 356:130-136(1994).
 CC -1- FUNCTION: May play a role in adhesive functions and in cell-cell
 CC interactions, metastasis and signaling. May provide a protective

Query Match 83.1%; Score 2130; DB 1; Length 1255;
 Best Local Similarity 39.0%; Pred. No. 3.3e-95;
 Matches 489; Conservative 1; Mismatches 5; Indels 760; Gaps 1;

QY 1 MTPGQSPFFLLTLTLVTVVGSCHASSTPGKEKTSATQRSSVPSSTKNAVSMTSSV 60
 DB 1 MTPGQSPFFLLTLTLVTVVGSCHASSTPGKEKTSATQRSSVPSSTKNAVSMTSSV 60
 QY 61 LSHSPGSGSSTQGDVTLAPATEPPASGSAATMGQDVSVVTRPALGSTTPADVTS 120
 DB 61 LSHSPGSGSSTQGDVTLAPATEPPASGSAATMGQDVSVVTRPALGSTTPADVTS 120
 QY 121 APDNKPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT----- 164
 DB 121 APDNKPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT----- 164
 QY 165 ----- 164
 DB 165 ----- 164
 QY 181 ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 240
 DB 181 ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 240
 QY 165 ----- 164
 DB 241 ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 300
 QY 165 ----- 164
 DB 301 ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 360
 QY 165 ----- 164
 DB 361 ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 420
 QY 165 ----- 164
 DB 421 ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 480
 QY 165 ----- 164
 DB 481 ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 540
 QY 165 ----- 164
 DB 541 ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 600
 QY 165 ----- 164
 DB 601 ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 660
 QY 165 ----- 164

```

Cc CC POLARIZED EPITHELIAL CELLS.
Cc CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF EPITHELIAL
Cc CC TISSUES. ABERRANTLY EXPRESSED IN EPITHELIAL CARCINOMAS.
Cc CC -1- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
Cc CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
Cc CC -----
Cc CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
Cc CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc CC the European Bioinformatics Institute. There are no restrictions on its
Cc CC use by non-profit institutions as long as its content is in no way
Cc CC modified and this statement is not removed. Usage by and for commercial
Cc CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
Cc CC or send an email to license@sib-sib.ch).
Cc CC -----
Cc CC EMBL; M64683; AAA39756.1; -
Cc CC EMBL; U16175; AAA98538.1; -
Cc CC EMBL; M65132; AAA39755.1; -
Cc CC EMBL; M64928; AAA39755.1; JOINED.
Cc CC EMBL; M77226; AAA39754.1; -
Cc CC PIR; A39344; A39344.
Cc CC MGI; MGI:97231; Muc1.
Cc CC InterPro; IPR000082; SEA_domain.
Cc CC Pfam; PF01390; SEA; 1.
Cc CC SMART; SM00200; SEA; 1.
Cc CC PROSITE; PS50024; SEA; 1.
Cc CC GlycoProtein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
Cc CC Repeat.
Cc CC -----
Cc CC CHAIN 1
Cc CC 1 20
Cc CC 21 630
Cc CC 21 535
Cc CC 536 556
Cc CC 557 630
Cc CC 42 366
Cc CC 42 61
Cc CC 62 81
Cc CC 82 101
Cc CC 102 122
Cc CC 123 143
Cc CC 144 164
Cc CC 165 184
Cc CC 185 204
Cc CC 205 225
Cc CC 226 246
Cc CC 247 256
Cc CC 257 286
Cc CC 287 306
Cc CC 307 326
Cc CC 327 346
Cc CC 347 366
Cc CC 366 376
Cc CC 376 386
Cc CC 386 396
Cc CC 396 406
Cc CC 406 416
Cc CC 416 426
Cc CC 426 436
Cc CC 436 446
Cc CC 446 456
Cc CC 456 466
Cc CC 466 476
Cc CC 476 486
Cc CC 486 496
Cc CC 496 506
Cc CC 506 516
Cc CC 516 526
Cc CC 526 536
Cc CC 536 546
Cc CC 546 556
Cc CC 556 566
Cc CC 566 576
Cc CC 576 586
Cc CC 586 596
Cc CC 596 606
Cc CC 606 616
Cc CC 616 626
Cc CC 626 630
Cc CC -----
Cc CC Query Match 46.9%; Score 1202.5; DB 1; Length 630;
Cc CC Best Local Similarity 45.0%; Pred. No. 3.9e-51;
Cc CC Matches 286; Conservative 63; Mismatches 141; Indels 145; Gaps 14;

```

```

QY 1 MPTGNSPFLILLTLVL-----TYTSGHAST-----PG 32
DB 1 MPTGIRAFPLILLTLVLSGLFALPSENVSTSSQOTSSSLASTTTPVHNSNDPATRP 60
QY 33 GKEKTSATQSSVPSSTERNVSTSSVLSGHS-----PGSGSTTQGGDTLAPATP 86
DB 61 GUSTSSPQSTSSPATRAPEDSTSTALSGTSSPATRAPVNSASSPVANGDTSATSP 120
QY 87 A--SGSAATWGQDVTSVPTRPALGSTTPPAHDVTSAPDNKPAGSTAPPAHGTSPATP 144
DB 121 LKDSNNSPVHSGTSSAATTAEDVSTSSPVHGGTSSPATSPGDSSTSSPHSTSSPAT 180
QY 145 RPPPGST-----APAHGVTSP 162
DB 181 RAPEDSTSTAVLSGTSSPATRAPVDSTSSPVADDTSSPATSLSDASSPVANGGTSSP 240
QY 163 DRRPAPGSTAPPAHG-----VTAPDNR-----PALGSTAPVHNTS-- 200
DB 241 ATSPAPDSTSSPVHSSASIQNIKTSDLASDPHNGTSTVTTSSLSGASPSPHSGTSTP 300
QY 201 -----ASGA-----SGASTVHNGTSP 220
DB 301 TMSGSVLTATTPVSSMPSTTKTSSGATIPDHNGSVLPTSSVLSGASVLYN--TSAT 359
QY 221 ATTPPASKTPTSPSHSDPTTLASHSTKTDASTHSTVPLTSSNHSSTPQSLSTV 280
DB 360 A--TTPVNSNGTOPSPVSPQVPSPTMAITSSHSTLASSSYSTVPSTPSSNS--SPQLSVGV 417
QY 281 SFFPLSFHISNLPNSLEDPTDYQELQDISEMFLQIYKQGGFLGSLNIXKRPSTV 340
DB 418 SFFPLFETIYQHPNSSLSDPSSNYYQELKRNISGLFLQIF--NEDFLGISIKKRSQV 476
QY 341 VOLTLAPFEGTINVDVETOFNQYKTEAASRYNLTSDVSAGHPFPPSAOSGAGVPG 400
DB 477 VESTVTFEKGFSASDVSKLQHKKEAD--YNLTJSEVNVNEMGPPSAQSRGVPGMG 535
QY 401 ILVLTVNVLALAIYVLAIAVCCCRKXNGOLDTPAPADYTHMSXPTVTHRRVYP 460
DB 536 IALVLVLCILVALAIYFLALANCCCRKXSGQDLDFPQDTHHMSYPTIRHGRVYP 595
QY 461 PSTDRSPYEKVSAGNGSSSYTPNPAVATSNAL 495
DB 596 PSTKRPQYEVESAGNGSSSYTPNPAVATSNAL 630

```

RESULT 4
MUC1 MESAU STANDARD; PRF: 676 AA.
AC 060528;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucin 1 precursor.
GN MUC1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euteleia; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
RX NCBI_TaxID=10036;
RN 11;
RP SEQUENCE FROM N.A.
RC TISSUE=Tracheal epithelium;
MEDLINE=9632618; PubMed=8703480;
RA Park H., Hyun S.W., Kim K.C.;
RT "Expression of MUC1 mucin gene by hamster tracheal surface epithelial
cells in primary culture."
RL Am. J. Respir. Cell Mol. Biol. 15:237-244(1996).
CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CYTOSKELETON (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -1- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

EMBL: U36918; AAB53965.1; -
InterPro: IPR000062; SEA_domain.
DR Pfam: PF01390; SEA; 1.
DR SMART: SM00200; SEA; 1.
DR PROSITE: PS50024; SEA; 1.
KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
Repeat.
FT SIGNAL 1 25
FT CHAIN 26 676
FT DOMAIN 26 582
FT TRANSMEM 583 603
FT DOMAIN 604 676
FT SIGNAL 458 573
FT CARBOHYD 291 323
FT CARBOHYD 323 350
FT CARBOHYD 350 380
FT CARBOHYD 400 433
FT CARBOHYD 413 435
FT CARBOHYD 435 479
FT CARBOHYD 479 496
FT CARBOHYD 536 536
SQ SEQUENCE 676 AA; 67616 MW; 95F479B6EC5C3844 CRC64;

Query Match 45.1%; Score 1156.5; DB 1; Length 676;
Best Local Similarity 40.6%; Pred. No. 6.6e-49;
Matches 276; Conservative 61; Mismatches 154; Indels 189; Gaps 12;

```

QY 1 MPTGNSPFLILLTLVL-----TYTSGHAST-----PG 32
DB 1 MPTGIRAFPLILLTLVLDVDPNSVALSDTSSSTLNTTPVHSGSAPATSSAVNDATTP 60
QY 1 MPTGIRAFPLILLTLVLDVDPNSVALSDTSSSTLNTTPVHSGSAPATSSAVNDATTP 60
DB 1 MPTGIRAFPLILLTLVLDVDPNSVALSDTSSSTLNTTPVHSGSAPATSSAVNDATTP 60
QY 25 GHASG-----TPGSEKTSATQSSVPSST-----EKNAVSMTSSVLSG---- 63
DB 61 GHSGSAPPTSSAVNSATTPGHSGSAPPTSSAVNSATTPVHSGSAPVTSVNSAVNSATTP 120
QY 64 -HSPGS-----GSSTTQGGDTVTLAPATEPAGSAAATWGQDVTSVPTRPALGSTTP 114
DB 121 VHSGSAPPTSSAVNSATTPVHSGSAPVTSVNSAVNSATTPVHSGSAPVTSVNSAVNSATTP 180
QY 115 AHDVTSAPDNKPAPGSTAPPAHGTSPADTPPPGSTAPPAHGTSPADTPPPGSTAP 174
DB 181 VHSGSAPPTSSAVNSATTPVHSGSAPVTSVNSAVNSATTPVHSGSAPVTSVNSAVNSATTP 240
QY 175 AHGVTSAADNRPALGSTAPVHNTS----- 201
DB 241 VHSGSAPPTSSAVNSATTPVHSGSAPPTSSAVNSATTPVHSGSAPVTSVNSAVNSATTP 300
QY 202 -----SGSAS----- 206
DB 301 VPPGSMQTTAISGSANMP IHNGSLVPTTSSALVPTTSAHSGASAMTSSSDSLATTP 360
QY 207 -----GSASTVHNGTSPARATTPPASKTPTSP 235
DB 361 IDSQTSISTTKPAPATTPVHNGSLVPTTSSVLSGATYLLIHNDISTMAATTPVANGTQSSVP 420
QY 236 SHHSDPTTLASHSTKTDASTHSTVPLTSSNHSSTPQSLSTGVSFPLSFHISNLP 295
DB 421 SRPPTTPPPAVNSNSTIALSTYSTALSPAFSSHA-APQVSVGVFPLSFHIMHNRN 479
QY 296 SSIQEDPTDYQELQDISEMFLQIYKQGGFLGSLNIXKRPSPVAVVQTLAPFREGTINVA 355
DB 480 SSIQEDPTDYQELQDISEMFLQIYKQGGFLGSLNIXKRPSPVAVVQTLAPFREGTINVA 538

```

QY 356 DVEFQNYKTEAASRYNLTISDVSVHVPFSSAGAGVPGMGIALLVLCVLAVALI 415
 Db EVKSOQLHQBHOE-AAEYVLAISKINVEQFPSSAQSPVPGMGIALLVLCVLAVALI 597
 QY 416 VYLAIALAVQCRRKRYGOLDIFPARDTYHPMSEVPTTHGRVYVPPSSTRASPEYKVSAG 475
 Db 598 VYLAIALAVQCRRKRYGOLDIFPIODSYHPMSEVPTTHGRVYVPPSSTRASPEYKVSAG 657
 QY 476 NGSSSLSTYNPAVATSANL 495
 Db 658 N-GSSSLSTYNPAVATSANL 676
 RESULT 5
 VGP3_EBV STANDARD; PRT; 907 AA.
 AC P03201; P03201; (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Envelope glycoprotein GP340 (Membrane antigen) (MA) [contains:
 Glycoprotein GP220].
 GN BLUF1.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 RA Tuffnell P.S., Barrell B.G.,
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 RT Nature 310:207-211(1988).
 CC -1- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CM2 RECEPTOR ON HUMAN
 CC B-CELLS.
 CC -1- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
 CC ENVELOPE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: V01555; CAA24854.1; --
 DR PIR: A03762; COBR21.
 DR PIR: A03763; COBR22.
 DR PIR: S33008; S33008.
 KW Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.
 FT CARBOHYD 87 47 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 169 169 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 229 229 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 378 378 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 386 386 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 497 497 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 533 533 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 547 547 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 589 589 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 610 610 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 624 624 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 656 656 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 683 683 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 701 701 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 735 735 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 746 746 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 755 755 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 780 780 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 815 815 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 858 858 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 888 888 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPPLIC 502 698 MISSING (IN GP220).
 SQ SEQUENCE 907 AA; 94431 MW; 0750141CBAC52C9 CRC64;
 Query Match 9.2%; Score 236.5; DB 1; Length 907;
 Best Local Similarity 23.1%; Pred. No. 9e-05;
 Matches 123; Conservative 62; Mismatches 194; Indels 153; Gaps 22;
 QY 22 TCGGASST-----PGGEKTSATORSSVPSTEKNAVSMVSSVSSPGSGSPT 73
 Db 444 TTGGLPSSTHVPTNLTAPAGTPTVST--ADVTSPTTAGTSGASPTVPSPSPMDNGTES 501
 QY 74 QGQDVT--LAAPTSPASGAAMTQGDVTSV-----VTRPLSGTTPPPAH 116
 Db 502 KADMTSSSPVPTPTPNATSPPTAVTPTPNATSPPTAVTPTPNATSPFLTKTSPSTA 561
 QY 117 DVTSAPD-NKAPSGTAPPAHGVTSAPD-TTRPPGSTAPPAHGVTSAPD-TTRPPGSTAP 173
 Db 562 VTPPTPNATSPFLTKTSPSTAATTPTPNATSPFLTKTSPSTAATTPTPNATSPFLTKTSP 621
 QY 174 PAHG-----VTSAPDNRPALGSTAPPVNAVTSAGSA-----SGSASTL-- 212
 Db 622 QANATHTLIGTSPTPVVTSQPKMATISAVTGO--NHITSSISMSLSRPSNPTLSPS 679
 QY 213 -VHNGTSARATTPA-----SKSTPPSIPSHSDPTTLASHTTDASTHST 261
 Db 680 TSDNSTSHMPLTFAHPTGGENITVTPASISTHVSSTSPAPRGTTQASGCGNSST 739
 QY 262 -----VPLTSSNHSTSPQSTGVSPFPLSPHISMLCPNSSL-----EDP 301
 Db 740 TKRGEVNVTKRPPQMATSPQASGQKTAVPVTSQCKANKSTGSKHTTGAGRTSTER 799
 QY 302 STDYQELORDISEMFLQIKQGGFLGSLNKKPRGVSVVQGLTAFRGGTINNVHDVTCR 361
 Db 800 TTDY-----GDSSTPRRYNATYILPST----- 824
 QY 362 NQYKTEASRYNLTISDVSVH--VFPFSSAGAGVPGM-GIALLVV-CVVALAIYV 417
 Db 825 ---SSKLRFRMTFTSPPTVPAQVVPVPTSQ---FRPSNLSMLVQMASLAVLTL 876
 QY 418 LIALAVQCRRKRYGOLDIFPARDTYHPMSEVPTTHGRVYVPPSSTRASPEY 469
 Db 877 LLMVADCAFR-----NLSTSHT---YTPPYDAETV 906
 RESULT 6
 TCNA_TRYCR STANDARD; PRT; 1162 AA.
 AC P23253;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sialidase (EC 3.2.1.16) (Neuraminidase) (NA) (Major surface antigen).
 GN TCNA.
 OS Trypanosoma cruzi.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

NCBI_Taxid=5693;
 [1] SEQUENCE FROM N.A.
 RP STRAIN-S1410 X-10/4;
 RC MEDLINE=91277609; PubMed=1711561;
 RA Pereira M.E.A., Mejia U.S., Ortega-Barria E., Matilevich D.,
 Prioli R.F.,
 RT "The Trypanosoma cruzi neuraminidase contains sequences similar to
 bacterial neuraminidases, with repeats of the low density lipoprotein
 receptor, and type III modules of fibronectin."
 RI J. Exp. Med. 174:179-191(1991).
 RL [2]
 RN SUBCELLULAR LOCATION.
 RP MEDLINE=91376547; PubMed=1896773;
 RA Prioli R.F., Mejia U.S., Aji T., Alkawa M., Pereira M.E.A.;
 RT "Trypanosoma cruzi: localization of neuraminidase on the surface of
 trypomastigotes."
 RL Trop. Med. Parasitol. 42:146-150(1991).
 CC -1- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
 CC PARASITE INVASION OF CELLS.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
 CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
 CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
 CC synthetic substrates.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POSSIBLE).
 CC -1- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM
 CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.
 CC -1- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
 CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA
 CC PROTEIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: CONTAINS 3 BNR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M61732; AAA30255.1; -
 DR PIR: JH0557; JH0557.
 DR InterPro: IPR002860; GH_BNR.
 DR Pfam: PF02012; BNR_2.
 KW Hydrolyase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
 KW Phosphorylation.
 FT DOMAIN 1 457 CYS-RICH.
 FT REPEAT 23 34 BNR 1.
 FT REPEAT 163 174 BNR 2.
 FT REPEAT 209 220 BNR 3.
 FT DOMAIN 458 588 FIBRONECTIN TYPE-III
 FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.
 FT CARBOHYD 342 342 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 1162 AA; 120032 MW; 0704922189706A40 CRC64;
 Query Match 9.1%; Score 234; DB 1; Length 1162;
 Best Local Similarity 28.7%; Pred. No. 0.00015;
 Matches 87; Conservative 34; Mismatches 134; Indels 48; Gaps 11;
 QY 16 TULVTGSGH-----ASTPGKEKTSATQSSVPSSTKNAWMTSSVLSSHS 65
 DB 707 TPSTPVDSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHG 766
 QY 66 PGSGSSITGQDVTIAPTEPASGSAAT---WQODYSVYTPALGSTTP---AND 117
 DB 767 TPSTPVDSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHG 826
 QY 118 VTSAPNKKPAPGSTAP---AHGVTAP-DTRPPGSTAPA---AHGVTAPDTPAPAG 169

DB 827 TPSTPVDSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHG 886
 QY 170 STAP-----AHGVTAPDNRPALGSTAPPYNTTASGSGSACTIV 213
 DB 887 TPSTPVDSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHG 944
 QY 214 HNGSARATTTTASGSPSTPSRPHNS--DTPITLASHST-KTPDASTHST-VPLITSSN 269
 DB 945 HSTSTPVDSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTPADSSA 1004
 QY 270 HST 272
 DB 1005 HST 1007
 RESULT 7
 ID AMYH_YEAST STANDARD; PRT; 1367 AA.
 AC P08640; P08068;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glucoamylase S1/S2 precursor (BC 3.2.1.3) (Glucan 1,4-alpha-
 DE glucosidase) (1,4-alpha-D-glucan glucosylhydrolase).
 GN STAI OR STAI2 OR MAL5 OR Y1R019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c / AB972;
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 RA Gentsile S., Hamlyn N., Horenell T.S., Hunt S., Jagels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
 RA Raftery M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
 RA MEDLINE=87194600; PubMed=106330;
 RA Yamashita T., Nakamura M., Fukui S.;
 RT "Gene fusion is a possible mechanism underlying the evolution of
 RT STAI".
 RL J. Bacteriol. 169:2142-2149(1987).
 RN [3]
 RP SEQUENCE OF 1-31 FROM N.A.
 RC STRAIN=SPX101-1C; PubMed=3141213;
 RX MEDLINE=89031230;
 RA Pardo J.M., Jerez F., Zalacain M., Clares M.G., Jimenez A.;
 RT "Similar short elements in the 5' regions of the STAI and SGA genes
 RT from Saccharomyces cerevisiae".
 RL FEBS Lett. 239:159-164(1988).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -1- SIMILARITY: TO S.POMBE SPBC215.13.
 CC -1- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z38061; CA86176.1; -
 DR EMBL: M16164; AAA35014.1; -
 DR EMBL: M16165; AAA35015.1; -
 DR EMBL: X13857; CAA32069.1; -
 DR PIR: B26877; B26877.

```

DR   PIR; A26877; A26877.
DR   PIR; S48478; S48478.
DR   SDD; S0001458; MUC1.
KW   Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW   signal; Multigene family.
FT   SIGNAL      1
FT   CHAIN       1
FT   DOMAIN      21
FT   DOMAIN      22 1367
FT   DOMAIN      210 1367
FT   CARBOHYD     817
FT   CARBOHYD     874 874
SQ   SEQUENCE    1367 AA; 136110 MW; 91C082DEDD61AA9D CRC64;
                                SER/GUANYLASE SL/S2.
                                SER/THR-RICH
                                N-LINKED (GLICNAC...) (POTENTIAL)
                                N-LINKED (GLICNAC...) (POTENTIAL)
                                N-LINKED (GLICNAC...) (POTENTIAL)

Query Match      9.18; Score 23.5; DB 1; Length 1367;
Best Local Similarity 28.98; Fred. No. 0.00019;
Matches 88; Conservative 45; Mismatch 123; Indels 49; Gaps 12.

OY 20 VYTGSGHASTPCGKEKRTATQNSQVP--SSTERNAVNTSSVLSHS--FGSGSTT 73
Db 344 VYSSSTSSSAVPPTPSSSTTSSSAAPVYSSSTSSSAAPTSSSSAPVPTPSSSTT 403
OY 74 QGQDVTLAPATPEASGSAATWG-ODVTSVPVTRPALGSTTPPADVTSAPDNKAP- 128
Db 404 ESSSAPVYSSSTSSSAAPTSTTSSSAAPVYSSSTSSSAAPTSTTSSSAAPVPTSS 463
OY 129 ----SSTAPPAHGVLT---SADTRPPPGSTAPAAHGVTSAPDTR----PAPSTAPAH 176
Db 464 STTSSSAAPVYSSSTSSSAAPVPTPSSSTTSSSAAPTSSSSAPVPTPSSSTTSS 522
OY 177 GVTSAPDTRPALST---APVNHVTSAGSA--GSGSALTVNHGTSARVTTTPASGS- 229
Db 523 --SSAPAPPTSSSTTSSSAAPVTSSTTSSSAAPVPTPSSSTTSSSTTPVYSSSTSS 580
OY 230 --TPPSISHSSTPTTLASHST-----KTDASSTHSITVPLTSSNHS-- 273
Db 581 VPTPSSSTSSSAAPVPTPSSSTTSSSAAPATPSSSTTSSSAAPVYSSSTSSSAAPT 640
OY 274 PQLST 278
Db 641 PSSST 645

RESULT 8
Y889 CAEBL STANDARD: PRT; 3178 AA.
AC 009624; 009625; 096904;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK945.9 in chromosome 11.
GN ZK945.9/ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OC NCBI_TaxID=6239;
[1]
SEQUENCE FROM N.A.
RP Durbin R.;
RA STRAIN=Bristol N2;
RC Wilkinson-Sproat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RP REVISIONS.
RA Durbin R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC 1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC 1- SIMILARITY: CONTRAINS 1 GPS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation in
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

[illegible]

AC Q07284; 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Envelope glycoprotein GP340 (Membrane antigen) (MA).
 GN BLUF1.
 OS Epstein-Barr virus (strain AG876) (Human herpesvirus 4), and
 OS Epstein-Barr virus (strain P3HR-1) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 NC NCB1_TaxID=62830, 82829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AG876, and P3HR-1;
 RC MEDLINE=9331716; PubMed=8393237;
 RA Lees J.F., Arrand J.E., Pepper S.V., Stewart J.P., Mackett M.,
 RA Arrand J.R.;
 RT "The Epstein-Barr virus candidate vaccine antigen gp340/220 is highly
 RT conserved between virus types A and B.";
 RL Virology 195:578-586(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P3HR-1;
 RA Klein K., Mueller-Lantusch N.;
 RL Submitted (OCT1992) to the EMBL/GenBank/DBJ databases
 CC -1- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
 CC B-CELLS.
 CC -1- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
 CC ENVELOPE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; L07922; AAA02783.1; -;
 DR EMBL; L07923; AAA02787.1; -;
 DR EMBL; X67776; CAA47986.1; -;
 KW Membrane; Glycoprotein; Antigen; Late protein.
 FT CARBOHYD 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 533 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 589 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 624 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 680 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 886 AA; 92388 MW; 4394F6130DECCABA CRC64;

Query Match 8.9%; Score 229; DB 1; Length 886;
 Best Local Similarity 24.2%; P-Id. No. 0.0002;
 Matches 136; Conservative 56; Mismatches 205; Indels 164; Gaps 26;

Qy 28 SSTPGKEKETATQRSS-----VPSSTKKNVSMSTSVLSHSPGS-G 69
 Db 370 SGTPSGGCGNISGAPASRRFDITVGGCTAGTKLILIRTRINATTTTKVLFSAQPSSTT 429
 Qy 70 SSTPGQDQVTLAPTE---PAS-----GSAATWGDVTSVPVTRPALSGTTPPAHDVT 119
 Db 430 TSPITNTTGFAPAPPTTTGGTPSSSTHVPTNLTPASTGPTVSTADVTSPTPACTTGSASPVT 489
 Qy 120 SAP---DN---KPAQGSTAPPAHGVTSAPD-----TRPPGSTAPPAHGVTSAPD-- 163
 Db 490 PPSPRDNGTESKAPDMTSPTSATVTPPTPNATSPTPAVTTPPNATSPFLT-GKTSPTSAV 548
 Qy 164 TRPAGSTAPPAHGVTSAPD-NRPAISGTA-----PHNVTSAS-GSAGSGSTLVH- 214
 Db 549 TTPPTNATSPPAVTTTPPAITPFLGKTSPTSATVTPPTNATSPVGETSPQANTTNT 608
 Qy 215 -NGTSAR-ATTTPASKSTPSPSPSHSDPTFTLASHSTK-----TDASSTHST 261
 Db 609 LGSTSTPTVTSPPKATSAVTTGQHNITSSSTSMELRPSISSETLSPTSNDSTSHM- 667
 Qy 262 VPLTSSNHSSTPGLSTGVSPFLSFHISNDQFNSLDEPSTDYQELQORDISEMPLQIY 321
 Db 668 --PLTSAHPGTGSENIQVTPASTGTH--HVSTSSPAPRGTT----- 706
 Qy 322 KQCGFLGSLNKKPRGSSVYVQL-----TLAFREGTIN----- 353
 Db 707 SQASGFGMSSTSTPGEVNVTKGTPPKATSPQAPSGOKTAVTVTSTGGRANSTTGAKI 766
 Qy 354 -----VHDETQFNQYKTEAARVNLTT---SPVSVSHVPPFF---SAQSGAG 395
 Db 767 TTGHGARTSTPPTDYGSDSTPRRRVNAATYLRPSTSSKLPRMFTSPPTTAQATVP 826
 Qy 396 VPGW-----GIALLVV-CVVALALIVYLALAVCCCRKNYGQDIFPARDTYHMSB 448
 Db 827 VPTSPGPFNSMLVYQMASLAVLTLVLLVMADCAFR-----N 867
 Qy 449 YPTVTHRGVYPPSSTDRSPY 469
 Db 868 LSTSHT---YTPPYDAETV 885

RESULT 10
 AGAL YEAST STANDARD; PRT: 725 AA.
 ID AGAL YEAST STANDARD; PRT: 725 AA.
 AC P32323;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE A-aggglutinin attachment subunit precursor.
 GN AGAL OR YNR044W OR N1431.
 OS *Saccharomyces cerevisiae* (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NC NCB1_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=9104412; PubMed=2072914;
 RA Roy A., Lu C.F., Marfakas D.L., Lipke P.N., Kurjan J.;
 RT "The AGAL product is involved in cell surface attachment of the
 RT *Saccharomyces cerevisiae* cell adhesion glycoprotein a-aggglutinin.";
 RL Mol. Cell. Biol. 11:4196-4206(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
 CC SUBUNIT. S. CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
 CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
 CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
 CC AGGREGATION DURING MATING.
 CC -1- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
 CC A CORE SUBUNIT.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR

CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M60590; AAA34382.1; -
CC EMBL; 271659; CAA96325.1; -
CC PIR; S17031; S17031.
CC PIR; A41258; A41258.
CC SGD; S0005327; AGA1.
CC Glycoprotein; Cell adhesion; signal; GPI-anchor; Repeat;
CC Pheromone response.
CC SIGNAL 1 22
CC CHAIN 23 725
CC DOMAIN 53 493
CC REPEAT 53 493
CC REPEAT 395 493
CC DOMAIN 182 307
CC REPEAT 182 188
CC REPEAT 189 195
CC REPEAT 196 202
CC REPEAT 203 209
CC REPEAT 210 216
CC REPEAT 217 223
CC REPEAT 224 230
CC REPEAT 231 237
CC REPEAT 238 244
CC REPEAT 245 251
CC REPEAT 252 258
CC REPEAT 259 265
CC REPEAT 266 272
CC REPEAT 273 279
CC REPEAT 280 286
CC REPEAT 287 293
CC REPEAT 294 300
CC REPEAT 301 307
CC SEQUENCE 725 AA; 73353 MW; 70420C851B0B01F8 CRC64;
Query Match 8.8%; Score 225; DB 1; Length 725;
Best Local Similarity 22.8%; Pred. No. 0.00026;
Matches 94; Conservative 80; Mismatches 152; Indels 86; Gaps 15;
QY 16 TVLVTVSGHASSSTPGGKETSATGRSSVPSSTKNAVSMTSSVLSHSPGSGSTTQG 75
DB 164 SLISPLYT-SLISSTSSNPPTTSLSTSTSPBST-STSPBSTSTSSSTSSSTSSSTSS 221
QY 76 OVVTIAPAPAPASGAATMGQDVTSVPTRPALSGTTPPAHDVTSAPDNKPAFGSTAPPA 135
DB 222 SSTSPSTSTSSSLTS-----TSSSSTSTSSGSET-----TSSSSTSPSTSS 270
QY 136 HGVTAPAPTR-PPPGSTAPPAHGVTSAP---DTPAPAGSTAPPAHGVTSAP-DNKPALGS 190
DB 271 SSTSPSTSPSKSTASSTSTSTSTSPSLTSSSTPLASTSPSTSTSTFTDSTSSGS 330
QY 191 T-----APVNAVTSAGS-----ASGASATLVHNGTSARATTPASKSTP 231
DB 331 SLASSSTSVLSYSDSTVSVPTSSNVAPESMSTSTVEFTVVSOGSSSYITKSISTTI 390
QY 232 PSIP-----SHSDPTTLASHST-KTDA-S-THHSTVPLPLTS 268
DB 391 PPSMSSTFTTVGVTTMYTTCPPYSSESESTTLTSHMETVTTDAVCHHSCHMSQSTTS 450
QY 269 NSTSPOLSTGVSFPLSFHISNLQFNLSLEDPSTDYVLCQDLIDEMVLCQYQGGFLG 328
DB 451 LITSSIKST-----KVAIVSTSTVBSSTACSTCAETSHSY----- 488

QY 329 LSNIXRPPGSVVVOLTAFREGTINVDVETQNOYKTEASRYNLTISPVS 380
DB 489 -SSVQVASSSVTQQTSTKSWSMWTSDDEPNKH---ATGKHVHTSSGTS 536
RESULT 11
VGLX_HSYEB STANDARD; PRT; 797 AA.
AC VGLX_HSYEB
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein X precursor.
GN 71.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316 (1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M8664; AAE02506.1; -
CC DR PIR; H36802; VGBEX1.
CC KW Glycoprotein; Transmembrane; Signal.
CC SIGNAL 1 22
CC CHAIN 23 797
CC DOMAIN 23 465
CC TRANSMEM 766 790
CC CARBOHYD 590 590
CC SEQUENCE 797 AA; 80342 MW; 50C9ED9211FE5B2 CRC64;
Query Match 8.6%; Score 221; DB 1; Length 797;
Best Local Similarity 26.3%; Pred. No. 0.00044;
Matches 103; Conservative 41; Mismatches 149; Indels 96; Gaps 14;
QY 11 LLLLVTLVTVTGS-----GHASSTPGGEKETSAP-----TQSSVPEST 49
DB 9 LLLCMVSVIYATGTTTTTETTTSSSSSTSGSGSTSSGTTNSSSPTTSSSPTTSS 68
QY 50 EKNAVMTS-----SVLSHSPGSGSTTQ-----GQVTLAP- 82
DB 69 HTSPSTSTGSSSTRAATSSSAPSTASSTTSPSTSTETTTTTPPASTTTTTPPA 128
QY 83 --ATPEPASGAATMGQDVTSVPTRPAL-GSTTPPAHDVTSAPDNKPAFGST-----A 132
DB 129 TAATTTAVTPAASATATTAATATSTPTTTPSTTTTATTTVPTTASTTTTDTTAA 188
QY 133 PPAHGVTSAPDTPPGSTAPPAHGVTSAPDTPPAFGSTAPPAHGVTSAPD----- 183
DB 189 TTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 248
QY 184 ---NRPAGSTAPRVHNVTSAGSAGSAS-----TVHNGTSARATTT---TPAKS 229
DB 249 TTAATTTAATTTGSGTSGSTSTTGASTSPASATATSTPTSTSAATTTSTPTTSA 308
QY 230 T-----PPSIPHSNDR-----TTLASHSTKTDASSTHSTVPLPLTSNHSSTPOLS 277
DB 309 TSAESTTEATPTPTTPTTTPSAATTAATTSPTSVASHTTSATTAATTAFTTSH-TSPDS 367
QY 278 TGVSFFPLSFHISNLQFNLSLEDPSTDY 306


```

FT CARBOHYD 4881 4881 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 4888 4888 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 4955 4955 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 4970 4970 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 5019 5019 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 5038 5038 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 5069 5069 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CONFLICT 1351 1351 H -> L (IN REF. 3)
FT CONFLICT 1412 1412 T -> S (IN REF. 3)
FT CONFLICT 1449 1449 L -> P (IN REF. 3)
FT CONFLICT 1504 1504 M -> T (IN REF. 3)
FT CONFLICT 1504 1504 G -> S (IN REF. 2)
FT CONFLICT 4192 4192
SQ SEQUENCE 5179 AA; 540295 MW; 85CD1571F89A5663 CRC64;

Query Match
Best Local Similarity 26.0%; Score 216.5; DB 1; Length 5179;
Matches 82; Conservative 46; Mismatches 126; Indels 61; Gaps 13;

QY 2 TGTGSPFLILUVTLVVTSGHASSTGGEKETSATQSSV----PESTKNAVNT 57
DB 1478 SEPTTP-----SPTTTPPTTPTTSPMTPTTPASTTTPPTTPTTPTT 1330
QY 58 SVLSHSRSGSS--TTCODVTLAPTEPAGSAATWGDVTSVPTPALGTT-- 112
DB 1531 TTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 1586
QY 113 -PPADVTASAP---DNKPAPGSTAPPAHGVTSAPTRP--PQGSTA-----PAA 155
DB 1587 SPTTTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1646
QY 156 HGVTSAPTRAPGST--APPAGVTSAP---DNKPALGSTAPPAHNTASASASGS 208
DB 1647 TTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 1706
QY 209 ASTLVHNGTSKATTPPASKTPPSIPSHSDPTPLASHTKDS--STHSVTPPLT 266
DB 1707 -----SPTTTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 1751
QY 267 SSMHSTFQSLSTGV 281
DB 1752 TSSPLTTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1766

RESULT 13
DAN4_YEAST STANDARD; PRT; 1161 AA.
ID DAN4_YEAST STANDARD; PRT; 1161 AA.
AC P47179;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell wall protein DAN4 precursor.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN 11
RA SEQUENCE FROM N.A.
RT Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RP REGULATION.
RX MEDLINE=2111166; PubMed=11160904;
RA Cohen B.D., Serflin O., Abramova N.E., Davies K.J., Lowry C.V.;
RT "Induction and repression of DAN1 and the family of anaerobic
RT mannoprotein genes in Saccharomyces cerevisiae occurs through a
RT complex array of regulatory sites."
RL Nucleic Acids Res. 29:799-808(2001).
CC -1- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SRP1 / TPI1 FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content, in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 249651; CA99684.1;
DR SCD: S0003812; DAN4;
DR Ites-Pro: IPR000992; SRP1_TPI1;
DR Pfam: PF00660; SRP1_TPI1;
DR PROSITE: PS00724; SRP1_TPI1;
KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
FT CHAIN 1
FT PROPEP 1147 1146
FT LEIDP 1146 1146
SQ SEQUENCE 1161 AA; 118358 MW; 7954C15069F0CA58 CRC64;

Query Match
Best Local Similarity 25.6%; Score 208.5; DB 1; Length 1161;
Matches 61; Conservative 46; Mismatches 148; Indels 41; Gaps 8;

QY 16 TVLVVTSGHASSTGGEKETSATQSSVPESTKNAVNTSVLSHSRSGSSSTTG 75
DB 123 TSTSTTKKSTSTTP---TPTTSTSTSTPTTSTSTSTSTSTSTSTSTSTST 178
QY 76 QDVTLAPTEPAGSAATWGDVTSVPTPALG--STTPPAHGVTSAPDNKPAGSTAP 133
DB 179 TPTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 236
QY 134 PAHGVTSAPDTPPEPSTAPAAHGVTSAPDTPAGSTAPPAHGVTSAPDNKPAGSTAP 193
DB 237 STTSQTSKSTSTPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 291
QY 194 PPAHNTASASASASTLVHNGTSAPATTPPASKTPPSIPSHSDTP-----TT 244
DB 292 TSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 349
QY 245 LASHSTKTD-----ASTHSTVPLPT-----SMHSTFQSLSTGVSPFLASF 287
DB 350 SATTTTSDTYISSSPSQVTSAPETVSEVTSVSEPTSSQVTSAPETVSEFTSSV 409
QY 288 HISNLQFNSSLDEPST 303
DB 410 EPTRSSQVTSAPETT 425

RESULT 14
C114_MOUSE STANDARD; PRT; 573 AA.
ID C114_MOUSE STANDARD; PRT; 573 AA.
AC P19467;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell surface antigen 114/A10 precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RA SEQUENCE FROM N.A.
RT MEDLINE=89197960; PubMed=2784793;
RA Dougherty G.J., Kay R.J., Humphries R.K.;
RT "Molecular cloning of 114/A10, a cell surface antigen containing
RT highly conserved repeated elements, which is expressed by murine
RT hemopoietic progenitor cells and interleukin-3-dependent cell
RT lines."
RL J. Biol. Chem. 264:6509-6514(1989).
CC -1- FUNCTION: MAY HAVE A POSITIVE REGULATORY ROLE IN THE CELLULAR

```

```

CC      RESIDUE TO IL-3.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- PTM: THE SER/THR-RICH TANDDEM REPEATS MAY SERVE AS SITES OF
CC      EXTENSIVE GLYCOSYLATION.
CC      -1- PTM: THE MOTIF SER-GLY CONSERVED IN ALL 8 SER/THR-RICH REPEATS MAY
CC      SERVE AS THE SITE OF ATTACHMENT OF GLYCOSAMINOGLYCAN SIDE CHAINS.
CC      -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC      -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.1sb-sdb.ch/announce/
CC      or send an email to license@1sb-sdb.ch).
CC      -----
DR      EMBL; J04634; AAA57239.1; -.
DR      PIR; A33533; A33533.
DR      MGd; MG1:103190; Ly64.
DR      InterPro; IPR000561; EGF-like.
DR      InterPro; IPR000082; SEA_domain.
DR      Pfam; PF01390; SEA; 1.
DR      SMART; SM00181; EGF; 1.
DR      SMART; SM00001; EGF-like; 1.
DR      SMART; PS00200; EGF; 1.
DR      PROSITE; PS00022; EGF_1; FALSE_NEG.
DR      PROSITE; PS01186; EGF_2; 2.
DR      PROSITE; PS50024; SEA; 1.
DR      GlycoProtein; Signal; Antigen; EGF-like domain; Repeat; Transmembrane.
KW      SIGNAL. 1 17 POTENTIAL.
FT      CHAIN * 18 573 CELL SURFACE ANTIGEN I14/I10.
FT      DOMAIN 18 480 EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 481 508 POTENTIAL.
FT      DOMAIN 509 573 CYTOPLASMIC (POTENTIAL).
FT      DOMAIN 17 230 8 X TANDDEM REPEATS. SER/THR-RICH.
FT      REPEAT 17 43 1.
FT      REPEAT 44 70 2.
FT      REPEAT 71 97 3.
FT      REPEAT 98 124 4.
FT      REPEAT 125 151 5.
FT      REPEAT 152 178 6.
FT      REPEAT 179 205 7 (NEAR PERFECT).
FT      REPEAT 206 230 8 (APPROXIMATE).
FT      DOMAIN 233 273 EGF-DIEX 1.
FT      DOMAIN 274 384 SEA.
FT      DOMAIN 385 425 EGF-DIEX 2.
FT      DOMAIN 425 467 EGF-DIEX 3.
FT      DISULFID 237 248 BY SIMILARITY.
FT      DISULFID 242 257 BY SIMILARITY.
FT      DISULFID 259 272 BY SIMILARITY.
FT      DISULFID 389 402 BY SIMILARITY.
FT      DISULFID 394 408 BY SIMILARITY.
FT      DISULFID 410 424 BY SIMILARITY.
FT      DISULFID 429 441 BY SIMILARITY.
FT      DISULFID 433 451 BY SIMILARITY.
FT      DISULFID 453 466 BY SIMILARITY.
FT      CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE 573 AA; 58701 MW; 1154C4F04E2D5A9 CRC64;
Query Match 8.14; Score 207; DB 1; Length 573;
Best Local Similarity 25.64; Pred. No. 0.0015;
Matches 101; Conservative 66; Mismatches 165; Indels 62; Gaps 18;
Ox 10 FILLLILLYLVYVTV-SGHASSTPGGEKETS-AIQRSSVPSSTBEKNAVSTSS----- 59
Db 4 FILLLSLILLYLVYVTVSSSSQSSSTSSSGGASPTTVQSSQSSSSQSSSTTSSSGASPT 63
Ox 60 VLSHSPSSGS--STTQGDVTLAPATPAEGASATWAGQDVTSVVPTRADVGSTTPPAHD 117
Db 64 TVQSSQSSPSSSSQSSSTTSSSGASPTTVQSSQSSPSSSSQSSST--TTSSSGASPT-T 118

```

Query Match 8.1%, Score 206.5, DB 1, Length 605,

Thu May 8 16:14:04 2003

us-09-658-621b-2.rsp

Page 14

Best Local Similarity 24.8%; Pred. No. 0.0016;
Matches 112; Conservative 66; Mismatches 179; Indels 95; Gaps 18

```

QY 33 GEMTSAIQRBSVSVSTREKNAVMSVSLSSHSRGSSTTQGGDVTIAPATEPASGAA 92
Db 111 GCPPLSSVQ--SVETSTESSVYSSSSITSS-----SSTSVVTTTISPLTSS---- 158
QY 93 TWCQVTSVPYTRPALGCTTPPADHVTSA--PDKKAPGSGTAPRAHGYTSAADTRPEOST 151
Db 159 -----TPTLT--TASTTSTSTDTISALPTTYSKULSTIPT--STTSSTSTTSTSS 208
QY 152 APAAAGVTSAPDTRAPAGSTAPRAHGYTSAADTRPALGSGTAPRHNVNTSASGASGAST 211
Db 209 TSTTVTSVSSSTSTTSTSTSTSTSTSTSTSSSSSTPTTSSAP-----I6STSTSTSTST 263
QY 212 LVHNGTSARATTPPASKSPPEIPESHSDPTTLASHSKTSDASTSHSTVDPPLTSSHS 271
Db 264 STTSTSTSSAAT--SSNTTPTSTTTTSSPSTABSTTIVYTSTASPTTSTTSSVQ 321
QY 272 TSPQLS--TGVSFFPLSPFHSUQPNSSLSDBSTDYQELORDISEMFTQIKQG--GFLG 328
Db 322 TSKTSVTVTSTVHTMDNISEI-----TSRYLTMKKVTT--QIYSSFLGAT 367
QY 329 LSNIRFSGSVVQOLTLAREEETINVDHVELOQNKTEASRNLITISDVSVSHVPPF 388
Db 368 TSVAVTTSASVSGEITNNNSNTTNSNTPNKSTEKKGWSPGKIAATPVVV----- 420
QY 389 SAQSGAGVPGMGIALVIVVCVVALAVITLALVAGQCRK-----NYGDLIPARDTY 443
Db 421 -----GVVCVVICIL-----IYLHHYKRRPKAKQDFENYDSKFY----- 458
QY 444 HPMSEYF-----TYTHGRVYPPSGSDRSPY 469
Db 459 --OSKRPNEVTTTTLHT-----PSPSNSSTF 482

```

Search completed: May 1, 2003, 23:10:54
Job time : 24.6786 secs

Thu May 8 16:14:04 2003

us-09-658-621b-2.rpr

Page 1

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:08:21; Search time 52.0536 Seconds

(without alignments)
914.184 Million cell updates/sec

Title: us-09-658-621b-2

Sequence: 1 MPTGTQSPFFELLTLTVLTV.....NGSSLSLYTNPAVATSNL 495

Scoring table: BLASTSUM62

Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

PIR 73:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2087	81.5	1344	1 A35175	mucin 1 precursor,
2	1599	62.4	1347	2 S10572	mucin 1 precursor,
3	1202.5	46.9	630	2 A39344	mucin 1 precursor,
4	1191	46.8	631	2 A39344	mucin 1 precursor,
5	1174	45.8	285	1 S88146	mucin 1 precursor,
6	564.5	22.0	236	2 A60533	tumor-associated a
7	265.5	10.4	1334	2 A50568	probable multi-dom
8	240	9.4	377	2 A48018	mucin 7 precursor,
9	237	9.3	1032	2 A34433	hypothetical prote
10	237	9.3	2232	2 A34433	hypothetical prote
11	235.5	9.2	907	1 Q08577	membrane antigen g
12	234	9.1	1162	2 Q08577	exo-alpha-stilidas
13	233.5	9.1	1367	1 S88478	glycan 1,4-alpha-g
14	233	9.1	796	2 A34433	hypothetical prote
15	233	9.1	3570	2 A50568	mucin MUC5B, trach
16	232.5	9.1	786	2 A50568	hypothetical prote
17	231.5	9.0	528	2 A34433	gastric mucin (C10
18	229	8.9	886	2 A50568	glycoprotein 350/2
19	225.5	8.8	886	2 A50568	proteohosphoglyca
20	225.5	8.8	825	2 A50568	hypothetical prote
21	225	8.8	725	2 A50568	mucin 5AC (clone J
22	225	8.8	725	2 A50568	a-agglutinin core
23	222	8.7	534	2 A50568	serine-rich protei
24	221	8.6	797	1 VGBK1	glycoprotein X pre
25	220.5	8.6	1952	2 A48814	hypothetical prote
26	219.5	8.6	1275	2 A33369	hypothetical prote
27	219	8.5	317	2 A53715	mucin (clone PCM-2
28	218.5	8.5	322	2 A53715	apomucin precursor
29	216.5	8.5	3020	2 A43932	mucin 2 precursor,

30	215.5	8.4	4776	2 E95206	cell wall surface
31	213	8.3	867	2 A45463	membrane glycoprot
32	212.5	8.3	866	2 A45462	membrane glycoprot
33	210	8.2	605	2 A33913	hypothetical prote
34	210	8.2	2187	2 A33913	nascent polypeptid
35	209	8.2	292	2 S24169	mucin 4, tracheal
36	209	8.2	660	2 A40067	chitinase IEC 3.2.
37	208.5	8.1	1161	2 A33523	probable membrane
38	207	8.1	573	2 A33523	cell surface glyco
39	206.5	8.1	507	2 A43768	antifreeze glycope
40	206.5	8.1	605	2 A48940	hypothetical prote
41	205	8.0	598	2 A39403	probable nucleopor
42	204.5	8.0	43	2 S35724	MUC1 protein - hum
43	201	7.8	354	2 A46740	microfilarial shea
44	200	7.8	610	2 A46740	mucin 4, tracheal
45	199.5	7.8	2761	2 A46740	hypothetical prote

ALIGNMENTS

RESULT 1
A35175
mucin 1 precursor, repetitive splice form A [validated] - human
A35175
N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; epistat
necrotic mucin; polymorphic epithelial mucin (PEM)
N:Contents: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor,
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000
C:Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S5
R:Rigdenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A:Title: Epistatin, a carcinoma-associated mucin, is generated by a polymorphic gene en
A:Reference number: A35175; MUID:90202794; PMID:2318825
A:Accession: A35175
A:Molecule type: mRNA
A:Residues: 1-952, 1033-1344 <LIG1>
A:Cross-references: GB:W32739; GB:J05288; NID:G182121; PIDN:AAA35804.1; PID:G182124; GB
A:Experimental source: splice form A
A:Note: GenBank entries HUMEPIS1A1 and HUMEPIS1A2 present only the amino- and carboxyl-e
A:Accession: B35175
A:Molecule type: mRNA
A:Residues: 1-19, 29-952, 1033-1344 <LIG2>
A:Cross-references: GB:W32739; GB:J05288; NID:G182126; PIDN:AAA35806.1; PID:G182129; GB
A:Experimental source: splice form B
A:Note: GenBank entries HUMEPIS1B1 and HUMEPIS1B2 present only the amino- and carboxyl-e
R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Pest, N.; Butche
r, B.L. Chem. 265, 15286-15293, 1990
A:Title: Molecular cloning and expression of a human tumor-associated polymorphic epithel
A:Reference number: A35886; MUID:90368715; PMID:1697589
A:Accession: A35886
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19, 29-952, 1033-1344 <GEN>
A:Cross-references: GB:J05581; NID:G18869; PIDN:AAA5976.1; PID:G188670
A:Note: GenBank entry HUMCMAB includes one copy of the tandemly repeated sequence
R:Plan, M.S.; Bacter, S.K.; Qi, W.N.; Metzger, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A:Reference number: A35887; MUID:90368716; PMID:2394722
A:Accession: A35887
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19, 29-1109, 5', 1111-1339, 'A', 1341-1344 <LAN>
A:Cross-references: GB:J05582; NID:G189598; PIDN:AAA60019.1; PID:G189599
A:Note: GenBank entry HUMPMNU contains four fewer copies of the tandemly repeated seq
R:Wreschner, D.H.; Hareven, M.; Tsataly, I.; Smorodinsky, N.; Horev, J.; Zaretzky,
E. J. Biochem. 189, 463-473, 1990
A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gene
A:Reference number: S10571; MUID:90276413; PMID:2351132
A:Accession: S10572
A:Molecule type: mRNA
A:Residues: 1-19, 29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1344 <WR>

A:Cross-references: EMBL:X52229; NID:g37053
R:Weascher, D.H.
Submitted to the EMBL Data Library, March 1990
A:Reference number: S40293
A:Accession: 540293
A:Molecule type: mRNA
A:Residues: 1-19-29-155, 'P', 157-1175, 'P', 177-182, 'A', 184-212, 1033-1037, 'A', 1039-1344 <MRNA>
A:Cross-references: EMBL:X52229; NID:g37053; PDB:CA36767.1; PDB:g37054
R:Abe, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated c-myc mRNA
A:Reference number: A36735; MUID:50088473; PMID:2597151
A:Accession: A36735
A:Molecule type: mRNA
A:Residues: 1-142 'O', 144-167 'O', 164-166 <ABE>
A:Cross-references: EMBL:M1823; NID:g181542; PDB:AAA35757.1; PDB:g181543
R:Minamata, Y.; Miyachi, T.; Hamano, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
Biochem. 112, 603-615, 1992
A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglutinin (PNA) binding site
A:Reference number: J00235; MUID:93123189; PMID:1478919
A:Accession: J00235
A:Molecule type: mRNA
A:Residues: 996-1011, 'ES', 1014-1017, 1018-1032, 'T', 1034-1037, 1038-1057 <MAS>
A:Experimental source: gastric carcinoma cell
R:Zitman-Blont, S.; Baruch, A.; Elroy-Stein, O.; Kedar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994
A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine h
A:Reference number: S51026; MUID:95080414; PMID:7988707
A:Accession: S51026
A:Contents: annotation
A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region are
C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
Partial repeats. The repeat shown is defined by SmaI nuclease sites.
C:Comment: Serine and threonine residues in the tandem repeat domain are extensively glycosylated
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C:Genetics:
A:Gene: GDB:MUC1; PUM
A:Cross-references: GDB:120705; OMIM:158340
A:Map position: 1q21-1q23
A:Intons: 20/1; 62/3; 1184/2; 1230/1; 1270/3; 1320/3
A:Superfamily: polymorphic epithelial mucin
C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphic
F:1-134/Product: mucin 1 precursor, splice form A #status predicted <PREA>
F:1-62/Region: mucin 1 amino-terminal non-repetitive
F:1-23/Domain: signal sequence #link PDB #status predicted <SIGA>
F:1-19-29-23/Domain: signal sequence #link PDB #status predicted <SIGA>
F:1-19-29-124/Product: mucin 1 precursor, splice form B #status predicted <SIGB>
F:1-19-29-212,1033-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
F:118-1017/Region: 20-residue repeats (GSAPAAHGVSAADTRPP)
F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F:145-1272/Domain: transmembrane #status predicted <TM>
F:1046-1064,1118,1144,1222/Binding site: carboxylate (Asn) (covalent) #status predicted
F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match	81.5%	Score 2087	DB 1	Length 1344
Best Local Similarity	88.0%	Pred. No. 4	4e-99	
Matches 412;	Conservative	7;	Mismatches 47;	Indels 2;
			Gaps 1;	

Cy 28 SSTPGGEKHSANQRSSVSTTEKAAVMTSSVSSHPSSGSGSTTGQODVTAAPRA 87
Db 879 STRAPAHGVTSADYTPRAGSTAPPAAGVTSAPTRAPAGSTAPRAPHG-VTSAPDTR 936
Cy 88 SSCSAATWQDVSVPTPALGSTTPPADYTSANDKPRAGSTAPPAAGVTSAPDTRP 147
Db 937 PGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSADYTPRAGSTAPPAAGVTSAPDTR 996
Cy 148 PGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSADNRPALGSTAPPAHANTSSAGS 207
Db 997 PGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSADNRPALGSTAPPAHANTSSAGS 1056
Cy 208 SASLTVANGTSARAATTPASKSTPPSPSHSHSDPTVTLAASHSTKDASTHSTVPLTS 267
Db 1057 SASLTVANGTSARAATTPASKSTPPSPSHSHSDPTVTLAASHSTKDASTHSTVPLTS 1116

QY	266	SNHSPQLSTVSPFPLSPHISNLQPNPNSLEDSDTDYQOELQORDISMFQIYKQGGFL	327
Db	1117	SNHSTSPQLSTVSPFPLSPHISNLQPNPNSLEDSDTDYQOELQORDISMFQIYKQGGFL	1176
QY	328	GLSNIKRPSPVAVVQLTAFREGTINVDHVTQGNQKTEAASRNLTISPSVSVHPPE	387
Db	1177	GLSNIKRPSPVAVVQLTAFREGTINVDHVTQGNQKTEAASRNLTISPSVSVHPPE	1236
QY	388	FSAGSAGVPGMGIMLTLVQVVALATYVILALVQCCRRKRYGQDDIPFADTYHPMS	447
Db	1237	FSAGSAGVPGMGIMLTLVQVVALATYVILALVQCCRRKRYGQDDIPFADTYHPMS	1236
QY	448	EYFVYTHGKRVPSSTDRSPYEKVSAGSGSSLSYTNPNVAATSANL	495
Db	1297	EYFVYTHGKRVPSSTDRSPYEKVSAGSGSSLSYTNPNVAATSANL	1344
RESULT 2			
	musn1	precursor, secreted epithelial tumor antigen splice form - human	
	S10571		
	N:contnabs:	musn1 secreted breast-cancer-associated splice form	
	C:species:	Homo sapiens (man)	
	C:date:	07-Apr-1994 #sequence revision 07-Apr-1994 #text_change 01-Dec-2000	
	C:accession:	S10571, JN0100, T56024, S09706, S10217	
	R:writer:	D. H.; Haeuwent, M.; Teartatay, I.; Smorodinsky, N.; Horev, J.; Zaretzky, J.	
	Eur J Biochem	189, 463-473, 1990	
	A:title:	Human epithelial tumor antigen CDNA sequences. Differential splicing may gener	
	A:reference number:	S10571, NMID:90276413; PMID:251132	
	A:accession:	S10571	
	A:molecule type:	mRNA	
	A:residues:	347, NRS>	
	A:cross-references:	EMBL:X5228; NID:936434; PIDN:CAA36477.1; PID:936435	
	R:writer:	J. Haeuwent, M.; Horev, J.; Zaretzky, J.; Weiss, M.; Udelich, J.M.; Garm	
	Gene 93	113,119, 1990	
	A:title:	isolation and characterization of an expressed hypervariable gene coding for a	
	A:reference number:	JN0100, NMID:91033045; PMID:1688329	
	A:accession:	JN0100	
	A:status:	preliminary	
	A:molecule type:	DN	
	A:residues:	'A',133,'Q',137-142,'F',144-153,204-208,'A',210-347<TSA>	
	A:cross-references:	GB:M5093, NID:9182252; PIDN:AAA59612.1; PID:9182253	
	R:writer:	P.K.; T.J.; Reynolds, K.; McDougall, P.O.; Purcell, D.F.J.; McKenzie,	
	J Mol Biol	142,13503-13505, 1989	
	A:title:	Reactivity of 156024 and 156024 human milk fat globulin antibodies with synthetic peptides.	
	A:reference number:	156024, NMID:89235154; PMID:2715633	
	A:accession:	156024	
	A:status:	preliminary; translated from GB/EMBL/DBBJ	
	A:molecule type:	mRNA	
	A:residues:	182-201<ERS>	
	A:cross-references:	GB:M26316; NID:9516622; PIDN:AAA36336.1; PID:9516623	
	R:writer:	S.D.B., 733-737, 1990	
	A:title:	Element of secondary structure in a human epithelial mucin core peptide fragm	
	A:reference number:	S09706, NMID:90253387; PMID:2339983	
	A:accession:	S09706	
	A:molecule type:	protein	
	A:residues:	182-201<TEN>	
	C:genetics:		
	A:gene:	GDB:MUC1, PUM	
	A:cross-references:	GDB:120705; OMIM:158340	
	A:map position:	1q21-q23	
	C:keywords:	alternative splicing; tandem repeat	
	F1-23/Domain:	signal sequence #status predicted <SIG>	
	F1-24-34/Product:	musn1, secreted epithelial tumor antigen splice form #status predict	
	F1-24-153,204-347/Product:	musn1, secreted breast-cancer-associated splice form #status	
Query Match			
Best Local Similarity	93.8%	Pred. No. 7, 5e-75;	
Match 315;	Conservative	0; Mismatch 1; Indels 20; Gaps 1;	
QY	1	MTPTGSPFLILITLVITVTGSGAASSTGAGKETSATQSSVPSSTKKNVMTSSV	60
Db	1	MTPTGSPFLILITLVITVTGSGAASSTGAGKETSATQSSVPSSTKKNVMTSSV	60

```

QY 61 LSSHPSGSGSTTGQGVTLAPATEPASGSAATWGDVTSVPTRPALGTTTPAHVTS 120
DB 61 LSSHPSGSGSTTGQGVTLAPATEPASGSAATWGDVTSVPTRPALGTTTPAHVTS 120
QY 121 APPDKPP-----PGSTAPAHGVTSAADTRPPPGSTAPAHGVTSA 160
DB 121 APPDKPPAHGVTSAADTRPPPGSTAPAHGVTSAADTRPPPGSTAPAHGVTSA 180
QY 161 APPTRPPPGSTAPAHGVTSAADTRPPPGSTAPAHGVTSAADTRPPPGSTAPAHGVTSA 220
DB 161 APPTRPPPGSTAPAHGVTSAADTRPPPGSTAPAHGVTSAADTRPPPGSTAPAHGVTSA 240
QY 221 ATTPPAKSTPSPSPSHSDPTTLASHSTKTDSSTHSTVPPPLTSSNHSSTPOLSTGV 280
DB 221 ATTPPAKSTPSPSPSHSDPTTLASHSTKTDSSTHSTVPPPLTSSNHSSTPOLSTGV 300
QY 281 SFFFLSFHISNLQFNSSLEDPSTDYQELORDISEM 316
DB 301 SFFFLSFHISNLQFNSSLEDPSTDYQELORDISEM 336

```

RESULT 3

```

A39344
tumor-associated mucin (MUC1) homolog precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 23-Feb-1997
A/Accession: A39344
R/Spicer: A.P.; Parry, G.; Patton, S.; Gendler, S.J.
J. Biol. Chem. 266, 15099-15109, 1991
A/Title: Molecular cloning and analysis of the mouse homologue of the tumor-associated m
A/loss of metastatic-like polymorphism.
A/Reference number: A39344; MUID:91332029; PMID:1714452
A/Accession: A39344
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-630 <SPI>
A/Cross-references: GB:M64928
C/Keywords: cytoskeleton; transmembrane protein

```

```

Query Match 46.9%; Score 1202.5; DB 2; Length 630;
Best Local Similarity 45.0%; Pred. No. 2.3e-54;
Matches 286; Conservative 63; Mismatches 141; Indels 145; Gaps 14;

```

```

QY 1 MTPGTQSPFFLLLLTLV-----TVVTGSGHAST-----PG 32
DB 1 MTPGIRAPFFLLLLLLSLKGFALPSEBNSVTSQDTSSSLASTTPVHSSNDPATRPP 60
QY 33 GEKETSATQKSVSTTEKNAVSMTSVLSHS-----PGSGSTTGQGVTLAPATEP 86
DB 61 GDSSTSPVQSTSSPATRAPBEDSTSTAVLSGTSPPATTAAPVNSASSPVAGHDTSSPATSP 120
QY 87 A--SGSAATMGQDVTSVPTRPALGTTTP--AHVTSAPDNKPAAGSTAPAHGVTSAAD 144
DB 121 LKDSNNSPVHSGTSSAPATTAAPVDSTSSPVHAGTSSPATSPGDSSTSSPHSSTSPAT 180
QY 145 TRPPGST-----APAHGVTSA 162
DB 181 RAPBEDSTSTAVLSGTSPPATTAAPVDSTSSPVHADDTSSPATSLSDSASSPVAGHGTSS 240
QY 163 DTRPAAGSTAPAHG-----VTSAPDN-----PALGSTAPVHANTVS-- 200
DB 241 PATSPTRDSTSSPVHSSASIONIKTSDLASTPDHNGSVTTSSALGATSPDHSGTST 300
QY 201 -----ASGSA-----SGSASTLVHNGTSA 220
DB 301 TNSSASVLAATTPVYSSMPSTTKVTSGSAILPDHNGSVLPTSSVLAGATSLVYN-TSA 359
QY 221 ATTPPAKSTPSPSPSHSDPTTLASHSTKTDSSTHSTVPPPLTSSNHSSTPOLSTG 280
DB 360 A-TTVNSGTPSPSPQVSVSTPATTSSTHSTVPPSTSSNS-SFQSLVGV 417
QY 281 SFFFLSFHISNLQFNSSLEDPSTDYQELORDISEMFLQIYKQGFGLSNIKFRPGSV 340

```

```

DB 418 SFFFLFFFIQHPHNSLSLEDPSSNYYDELKXNISGLEIF-NGFLLISIKRSSSV 476
QY 341 VOLTLAPREGTINHVETQFNQYKTEAASRYLTI SDVSVHPFPFSAQSGAGVPGNG 400
DB 477 VESTVFPREGTFSADVKSQLIQHKEDS-YNLTISEVKVEMQFPFSAQSRGQVGMG 535
QY 401 IALLVLCVVALATVYLLALAVCCCRKNGVGLDIFPARDTYHMSSEPTVTHGRVYP 460
DB 536 IALLVLCVVALATVYLLALAVCCCRKKSQGLDIFPDQTYHMSSEPTVTHGRVYP 595
QY 461 PSTDRSPYEVNSAGNGSSLSYTPAYAAATSA 495
DB 596 PSTDRQPYEVNSAGNGSSLSYTPAYAAATSA 630

```

RESULT 4

```

152257
epitailin - mouse
C/Species: Mus musculus (house mouse)
C/Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
A/Accession: 152257; 165210
R/Vos: H.L.; De Vries, Y.; Hilkens, J.
Biochem. Biophys. Res. Commun. 181, 121-130, 1991
A/Title: The mouse epitailin (Muc1) gene and its promoter. Rapid evolution of the repet
A/Reference number: 152257; MUID:92068178; PMID:1958179
A/Accession: 152257
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-631 <RBS>
A/Cross-references: GB:M77226; NID:g199835; PIDN:AAA39754.1; PID:g199837
A/Accession: 165210
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-631 <R2>
A/Cross-references: GB:M64683; NID:g199842; PIDN:AAA39756.1; PID:g199843
C/Genes:
A/Genes: Muc1
A/Introns: 20/1; 454/3; 472/2; 517/1; 557/3; 607/3

```

```

Query Match 46.9%; Score 1201; DB 2; Length 631;
Best Local Similarity 45.3%; Pred. No. 2.7e-54;
Matches 288; Conservative 63; Mismatches 139; Indels 146; Gaps 15;

```

```

QY 1 MTPGTQSPFFLLLLTLV-----TVVTGSGHAST-----PG 32
DB 1 MTPGIRAPFFLLLLLLSLKGFALPSEBNSVTSQDTSSSLASTTPVHSSNDPATRPP 60
QY 33 GEKETSATQKSVSTTEKNAVSMTSVLSHS-----PGSGSTTGQGVTLAPATEP 86
DB 61 GDSSTSPVQSTSSPATRAPBEDSTSTAVLSGTSPPATTAAPVNSASSPVAGHDTSSPATSL 120
QY 87 A--SGSAATMGQDVTSVPTRPALGTTTP--AHVTSAPDNKPAAGSTAPAHGVTSAAD 143
DB 121 SKDSNNSPVHSGTSSAPATTAAPVDSTSSPVHAGTSSPATSPGDSSTSSPHSSTSPA 180
QY 144 TRPPGST-----APAHGVTSA 161
DB 181 TRAPBEDSTSTAVLSGTSPPATTAAPVDSTSSPVHADDTSSPATSLSDSASSPVAGHGTSS 240
QY 162 DTRPAAGSTAPAHG-----VTSAPDN-----PALGSTAPVHANTVS-- 200
DB 241 PATSPTRDSTSSPVHSSASIONIKTSDLASTPDHNGSVTTSSALGATSPDHSGTST 300
QY 201 -----ASGSA-----SGSASTLVHNGTSA 219
DB 301 TNSSASVLAATTPVYSSMPSTTKVTSGSAILPDHNGSVLPTSSVLAGATSLVYN-TSA 359
QY 220 RATTTPPAKSTPSPSPSHSDPTTLASHSTKTDSSTHSTVPPPLTSSNHSSTPOLSTG 279
DB 360 IA-TTVNSGTPSPSPQVSVSTPATTSSTHSTVPPSTSSNS-SFQSLVGV 417
QY 280 VSFFLSFHISNLQFNSSLEDPSTDYQELORDISEMFLQIYKQGFGLSNIKFRPGSV 339

```



```

Query Match      10.4%; Score 265.5; DB 2; Length 1334;
Best Local Similarity 33.7%; Pred. No. 2.9e-06;
Matches 91; Conservative 14; Mismatches 90; Indels 75; Gaps 12;

QY 31 PGEKETATQSSVPSSTKNAVSMITSSVLSSHPGSGSSTTGGQDVTLPATPEPASGS 90
Db 237 PGEKATLHALLS-PSPPTPGSRPTGPTGPGPAGCAAGASGTDV-----ASGA 287
QY 91 AATWGDVTSVVTREPAL-----GSTTPPADVTAPDNKPAFG- 129
Db 288 GAASGPDPAAGSAPVAPGSGGAPGMPAPGTAPGSGSTAPPDHTASADTAPAGP 347
QY 130 ----STAPPAHGVTSAPDTR-PPGS-----TAPPAHGVTSAPDTRPAGSTAP- 173
Db 348 TSAPGTAAPAGTAAPAGTAPGAPGAPGTAAPVA-GTTAPGTAAPAGTAPGAPDTSY 406
QY 174 ----PAHGVTSAPDNRPALGSTAPVHNVTASGSA-----SGSASTLVH 214
Db 407 AAGTAAPVAGTTPAPGTAPAPGST-PAAGTVPAPGTAPAPGAPDGRBPVTGPAS- 460
QY 215 NCTSAATTP-ASKSTPSIFSHTSDTP 242
Db 461 -GTGPGATPPPEAAASASGAPSPAPBGP 489

RESULT 8
A48018
mucin 7 precursor, salivary - human
N:Alternate names: mucin, MG2; mucin, MG2a-T1; mucin, MG2a-T2; mucin, MG2b-T2
C:Species: Homo sapiens (man)
C>Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 07-May-1999
C:Accession: A48018; S29115; S29116; S29114
J:Bobek, L.A.; Tsai, H.; Biesbrock, A.R.; Levine, M.J.
R:Biochem. J. 287, 639-643, 1992
A:Title: Molecular cloning, sequence, and specificity of expression of the gene encoding
A:Reference number: A48018; PMID:9338636; PMID:7690757
A:Accession: A48018
A:Molecule type: mRNA
A:Residues: 1-377 <BOB>
A:Cross-references: GB:L13283
A:Experimental source: submandibular gland
A>Note: sequence extracted from NCBI backbone (NCBIN:137719, NCBI:P:137720)
R:Reddy, M.S.; Bobek, L.A.; Haraszthy, G.G.; Biesbrock, A.R.; Levine, M.J.
Biochem. J. 287, 639-643, 1992
A:Title: Structural features of the low-molecular-mass human salivary mucin.
A:Reference number: S29114; PMID:9307506; PMID:1445223
A:Accession: S29115
A:Molecule type: mRNA
A:Residues: 143-168 <RED>
A:Accession: S29116
A:Molecule type: protein
A:Residues: S', 71-79, 'N', 81-86, 'X', 89, 'X', 91, 'P' <RE2>
A:Molecule type: protein
A:Residues: 143-145, 'X', 147, 'XXX', 151-152, 'X', 154-158, 'X', 160-161, 'A', 163-164, 'X', 167-1
C:Genetics:
A:Gene: GDB:MUC7
A:Cross-references: GDB:138799; OMIM:158375
A:Map position: 4q33-4q31
A:Keywords: glycoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-377/Product: mucin 7, salivary #status predicted <AMT>
F:97, 128, 135, 146, 312/Binding site: carbohydrate (Aam) (covalent) #status predicted

Query Match      9.4%; Score 240; DB 2; Length 377;
Best Local Similarity 30.5%; Pred. No. 1.5e-05;
Matches 90; Conservative 44; Mismatches 91; Indels 70; Gaps 17;

QY 37 TSATQSSVPSSTKNAVSMITSSVLSSHPGSGSSTTGGQDVTLPATPEPASGSAATGQ 96
Db 118 SASRTKITLPNT--FLPQWATLISRMNVNTSSV--ATLAPVNSPAP-----Q 163
QY 97 DVTSVTRPALGSTTPPADVTAPDNKPAAGST--APPAHGVTSAPDTRPPGSTAPA 154

```

```

Db 164 DTTAP--PTPEATTP-----APSSAPPETTAP-----TPSATPAPSSAP 209
QY 155 AHGVTSAPDTRPA-----PGTAPPAHGVTSAPDNRA-----LGTAPVHNVTASGS 204
Db 210 E--TTAAPPTTPPTTAPPSAPPE--TTAAPPTSATTPAPLSSAP--ETTAVP 261
QY 205 ASGSASTLVHNGTSARATTT-----PASKSTPSIPSHSDT-----PTTLASH 248
Db 262 PTPSATLTPDSSASAPPETTAPPTPSATTPAPPSAPPETTAPPTTPSSPTLAP 321
QY 249 STKTDASGTHSTVPLTSSNHSIPOLSTG---VGFPLSPHNLQFNSLED 300
Db 322 TSETSAAPHOTTSTVTTOTTTTKOPTSAPQNKISFLL--YMKL-LNRIID 373

RESULT 9
T34433
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34433
R:Geisell, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34433
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1032 <GEI>
A:Cross-references: EMBL:U08046; PIDN:AACT0889.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2; 9'

Query Match      9.3%; Score 237; DB 2; Length 1032;
Best Local Similarity 30.5%; Pred. No. 6.2e-05;
Matches 103; Conservative 42; Mismatches 131; Indels 62; Gaps 14;

QY 16 TVLVYVGGG--HASTP-----GGEKETATQSSVPSSTKNAVSMITSSVLSSHPG 67
Db 357 STVTVVGGSSSTPASTPTIASSSSGSTVTVVAPGSSSTVGSSTPSSSSSGTWTSGS 416
QY 68 SGGSTTGGQDVTLPATPEPASG-----SAATWGDVTSVVTREPALGSTTPPADVTSA 121
Db 417 TGST-----VTAPVSSSTPSSSTPTIASSSSGSTVTVVSSSSSTTSSSPSSSAG 470
QY 122 PDKRA-----PGSTAPPAHGVTSAPDTRPPGST--APPAHGVTSAP--DTRPA 167
Db 471 TASTISGTSSTATTIVPGSSSVGSSQTQSAPSPSGTMTVSGPTGSTTVVPGSSTPA 530
QY 166 PGSTAPPAHGVTSAPDNRPALGSTAPVHNV--TSASGASGASTLVHNGTSARATTP 225
Db 531 PSSSNPSSSPASTGTTITGSSGSLIVTVSGSTVSGSTGTSQSTL-----ASSTAP 584
QY 226 ASKSTPSI-----PSHSDPTTTLASHTKTDASSTHSTVPL--TSSNHSI----- 272
Db 585 GSSSTVSSSSPPSSQSPAPVTSSTTPSGTSSQSPSPMNSSTPTGSSQSTTPSS 644
QY 273 ---SPQSTGVSPFLSFH--SNLQFNSLEDPESTD 304
Db 645 TNSPTGSGT-STFSVATEVTSQSTVPSGSLGTQSTN 681

RESULT 10
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisell, C.; Gattung, S.

```


QY	66	PGGSGSTTQGDVTLAATAPATEPAGSAAAT----	WGDVTSVVTREPALGSTTP-----	AHD	117
Db	767	TETPADSDSAHSTSTPVDSDSAHGTSTPVDSDSAHSTSTPVDSDSAHGTPTPVDSDSHS			826
QY	118	VTSPADNPAPAGSTAPP-----	AHGVTSAP-DTRPPPGSTAPA--	AHGVTSAPDTRPAPG	169
Db	827	TETPTVDSDSAHGTSTPVDSDSAHSTSTPVDSDSAHSTPTPVDSDSAHGTPTPVDSDSHS			886
QY	170	STAPP-----	AHGVTSAPDNPALGSTAPPVNNVTSAGSAGSASLTLY		213
Db	687	TETPTPADSDSAHSTSTPVDSDSAHSTSTPVDSDSAHGTSTPVD--	SSAHGTSTPTPADSDSA		944
QY	214	HNGTSAAKATTPPAKSTPPSPSPSHS--	DTPTTLASHST-KTDSASTHST--	VPELTGSN	269
Db	945	HSTPTSPPADSDSAHSTPTPADSDSAHSTPTPVDSDSAHSTPTPADSDSAHSTPTPADSDSA			1004
QY	270	HST	272		
Db	1005	HST	1007		

RESULT 13
548478
Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)
N/Alternative names: extracellular glucan 10-Sep-1999 #text_change 12-Nov-1999
C/Species: *Saccharomyces cerevisiae*
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999
A/Accession: 548478; A26877; B26877; S27281; J06123
R/Rowley, K.
submitted to the EMBL Data Library, October 1994
A/Reference number: 548478
A/Accession: 548478
A/Molecule type: DNA
A/Residues: 1-1367 <ROM>
A/Cross-references: GB:I247047; EMBL:Z38061; NID:9603997; PID:9763364; GSPDB:GNO0009; MIP
J.Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A/Title: Gene function is a possible mechanism underlying the evolution of STAL.
A/Reference number: A91831; MUID:87194600; PMID:3106330
A/Accession: A26877
A/Molecule type: DNA
A/Residues: 1-242 <YAM>
A/Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A/Accession: B26877
A/Molecule type: DNA
A/Residues: 762-1331 <YAZ>
A/Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R.Pardo, J.M.; Jance, E.; Zalacain, M.; Clatos, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A/Title: Similar short elements in the 5' regions of the STA2 and SGA genes from *Sacchar*
A/Reference number: S27281; MUID:89031230; PMID:3141213
A/Accession: S27281
A/Molecule type: DNA
A/Residues: 1-31 <PAR>
A/Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R.Lambrecht, M.G.; Bauer, F.F.; Marmur, J.; Peticola, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A/Title: Muc1, a mucin-like protein that is regulated by Msa10, is critical for pseudoph
A/Reference number: J06123; MUID:9632337; PMID:8710886
A/Accession: J06123
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-1367 <LAM>
A/Cross-references: GB:030626; NID:g1304386; PIDN:AA49609.1; PID:g1304387
C/Genetics:
A/Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A/Cross-references: MIPS:YIR019c; SGD:S0001458
A/Map position: 9R
C/Superfamily: yeast glucan 1,4-alpha-glucosidase homolog
C/Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
P:5-21/Domain: transmembrane #status predicted <TM1>
P:1350-1366/Domain: transmembrane #status predicted <TM2>

```

Query MatchScore      9.1%; Score 233.5; DB 1; Length 1367;
Best Local Similarity 28.9%; Pred. No. 0.00013;
Matches "88; Conservative 45; Mismatches 123; Indels 49; Gaps 12;

QY    20 VYTGGGHAATPTGGEKATKTORSVSP---SSTEKNAMVTSTSVLSHS---PGSGSTT 73
Db    344 VTISTSTSSSSAPVTPPSSSTTSSSSAPVTSSSSAPVTSSTSSAPVTSSTSSSAPVPVTPSSSTT 403

QY    74 QGDGYTLALPATPEPASGSAAITWG-QDVTSVTPVRPALGSTTPPAHDVTSAPDNKPAP--- 128
Db    404 ESSSAPVTSSTTSSSSAPVTSSTTSSSSAPVTSSTTSSSSAPVTSSTTSSSAPVTPSS 463

QY    129 ----GSTAPPAHGVY--SAPDRPPPGSTPAHAGVTSAPDR-----PAGSGTAPPAH 176
Db    464 STTSSSSAPVTSSTTSSSSAPVTPPSSSTTSSSSAPVTSSTTSSSSAPVTPPSSSTTES- 522

QY    177 GTVSAPDNRPALGST--APPVHNVTASAGSA--SSASATLVHNGTSARATTTPACKS- 229
Db    523 --SSAAPAPPTSSSTTSSSSAPVTSSTTSSSSAPVTPPSSSTTSSSTTPVTSSTSSSAP 560

QY    230 --TPSISPSHSDPTPLASHST-----KITASGTSHSIVPPLTSGNHSITS----- 273
Db    581 VTPPSSSTTSSSAPVTPPSSSTTSSSSAPAPTSSTTSSSAPVTSSTTSSSAPVTSSTTSSSAPVPT 640

QY    274 POLST 278
Db    641 PSSST 645

```

```

RESULT 14
T21460
hypothetical protein ZK945.10 - Caenorhabditis elegans
C|Species: Caenorhabditis elegans
C|Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C|Accession: T21460
R|Wilkinson, J.
submitted to the EMBL Data Library, March 1995
A|Reference number: Z19425
A|Accession: T21460
A|Status: preliminary; translated from GB/EMBL/DDBJ
A|Molecule type: DNA
A|Residues: 1-795 <M1>
A|Cross-references: EMBL:Z48582; PIDN:CAA8469.1; GSDDB:GN00020; CESP:ZK945.10
A|Experimental source: clone F2/B5
C|Genetics:
A|Gene: CESP:ZK945.10
A|Map position: 2
A|Intons: 62/3; 124/3; 169/1; 216/1; 262/3; 371/1; 581/3; 608/1

Query Match          9.1%   Score 233; DB 2; Length 796;
Best Local Similarity 26.8%, Pred. No. 7.5e-05;
Matches 115; Conservative 54; Mismatches 172; Indels 88; Gaps 19

Oy      19  TVVTGSGASSTPGCE--KETSAIGRSSVPSETEKNVASMISVSLSHSPCGS-----70
Db      302  TVTTAMTSISTSIPTSIITISTSTFTPTSTASTASTSTTGQSSSTITSPESTILSTSI 361
Oy      71  STTGQGDVTLAPATEPAGSAAWTQDVTSVVTPRPALGSTTPPADVTSADNKPPAGS 130
Db      362  PTTTTEIETISLTSSLPDNAICSLDETSTTTFITMLTSTT-----TEEP-----S 408

Oy      131  TAPPAGVTSAPDTRPPPGSTAPAAAGVTSAPDTRPAPAGSTAAPPAAGVTSAPDNRPALGS 190
Db      409  TSPTTTEVTSTST--VTTEPTTLLTSTAST-----STTEESTSTVTSTSP-----S 454

Oy      191  TAIPVNVYTSAGSGASGASATLVHNSTASARV-TTPAS-----KSTPSI-----PSGHSDT 241
Db      455  TSPVSTVTVSSS-----SSSTVVTPTSTESTSTSPSSVTVTSTTAIPSTSTGGPSSSST 509

Oy      242  PTTLASHRTKDASTAHST---VPLTSSNNHSNGPOLSGVGFPL-----SPHS---290
Db      510  PSSSTASSVSSSTASSSTGOSTSQSSSTTKSSTTSSGDTNPDPFVEAKATTTFIDSTSV 569

```

QY 291 NLOPNSL-----EDPSTDIYQELORDISMEPLQ-----IKGGGFLGAS 330
Db 570 NLTNLSGLIIGVQTSICTSPSTSNVSTTKD-GACFKYSVMRLGITYPASTFVGP 628
QY 331 NIKFRGSAV-----VOLTLA---FREGTINHDVETQFOQYTEAASRYNLTISDVASH 383
Db 629 NYTFRATMTDCKKYYIYANVYIQEYSSSTIESBSTSAVASSTSPSPSTLSTST 688
QY 384 VPPFPAQS 392
Db 689 VTEPSTRS 697

RESULT 15

T45025
mucin MUC5B, tracheobronchial (imported) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence, revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45025
R:Desseyn, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A:Reference number: 222899; MUID:9716151; PMID:9013550
A:Accession: T45025
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3570 <DES>
A:Cross-references: EMBL:272496; NID:g1834502; PID:CA96577.1; PID:g1834503
A:Experimental source: placenta
C:Genetics:
A:Gene: MUC5B

Query Match 9.1%; Score 233; DB 2; Length 3570;
Best Local Similarity 23.6%; Pred. No. 0.00038;
Matches 160; Conservative 50; Mismatches 215; Indels 252; Gaps 31;

QY 2 TPCTQSPFLILLTLVTVVGS---GHASTPGR---KETSATORSSVP-----SS 48
Db 1775 TPCTT---WILTELTATTTAALPHGTSPSTPGTWTLEPSTTATVTVTGSTATASS 1831
QY 49 TEKNA-----VMTSSVLSH-SFGSSSTTQGDVTLAPATEPASGSAATWGQDT 99
Db 1832 TRATAGTLLKVLSTVATPTVSSRATPSSPGATAPALPRLSTATTPTATS-----VT 1884
QY 100 SVPTRPALG-----STTPAHDVTSAPDNKP-----APG 129
Db 1885 AIRPS--SLGTWTRLGQTTTPATWSTATPSTPETVHTSVLTTTATTTTTRGSVATPS 1942
QY 130 STAPPAH-----GVTSAPDTRP-----P-----GSTAPAAHGVTSAPDTRP----- 166
Db 1943 STPGTATHTTKVPTTTTTFIATPSSSPGATLTPPWISTTTPTTGTGSTVTPSSIPGTH 2002
QY 167 -----APGSTAP-----AHGVTSAPDNRPALGST-A 192
Db 2003 TATVLTATTTTVAATGSMATPSSSTQSTGTPPLTATTATTATGSTMPSSTP--GTTPI 2060
QY 193 PPHANTSSAGSAGSAST-----LVHNTSARATTPASKSTPSP-- 235
Db 2061 PPVLTATTATPATSTVTPSSALGTTTPPVNTATTGRLSPSSPHTVPTAMTSAT 2120
QY 236 -----SHSDPTTTLASHSTKTDASTHSTVPLTSSNH-----STSPOLS 277
Db 2121 SGILGTHITEPSTSTSTPATGTTQPSST--PALSPHPSRTSPSPGTTTGH 2178
QY 278 TGVSFFELS-----PHINLOFNSSLEDPSTDYQ---ELORDISEMFLQIYKO-----G 324
Db 2179 RGTSTRTATATPSTKRTISTLLPSSPTAPITTVTTGCPQCAWSEMLDYSYMPGPSG 2238
QY 325 GFLGSLNINIKFRGSAVVVQ--LTLAFFEGTINHDVETQFOQYTEAASRYNLTISDVASH 383
Db 2239 DFTYTSNIRAAAGAVCEQPLGLECR----- 2263

QY 384 VPPFPAQSAGVPGKGIALLV-----LYC-----VLVALATVTLAIVQCRRKN 430
Db 2264 -----AAQAPGVPLAELEQVVECSLDPLGLVCNRNBOYKPFKCNRYEIRVFC-----N 2312
QY 431 YGOLDIPARDTYHMPWSEYP-----TYHTGARYVPPSSTDRSPFEKVSAGN 476
Db 2313 YGHCPSTPATSTATPSSSTPGTWTILTKLTATTATTETSTGSTATPSSSTQGP-----AGT 2367
QY 477 GGSSLSTYTPAVALTSA 493
Db 2368 PHVSTATTPTVTSKA 2384

Search completed: May 1, 2003, 23:13:51
Job time : 60.0536 secs

Thu May 8 16:14:07 2003

us-09-658-621b-26.closed.rat

Page 1

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:21:52 ; Search time 28 Seconds
(without alignments)
9.457 Million cell updates/sec

Title: US-09-658-621B-26
Perfect score: 44
Sequence: 1 ISEMPLQIY 9

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: * 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 66399

Minimum DB seq length: 9
Maximum DB seq length: 9

Post-processing: Minimum March 0%
Maximum March 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgm2_6/prodata/1/iaa/5A COMB pep:*
2: /cgm2_6/prodata/1/iaa/5B COMB pep:*
3: /cgm2_6/prodata/1/iaa/6A COMB pep:*
4: /cgm2_6/prodata/1/iaa/6B COMB pep:*
5: /cgm2_6/prodata/1/iaa/PCITUS COMB pep:*
6: /cgm2_6/prodata/1/iaa/Backfile1 pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	47.7	9	1	US-08-338-634-10 Sequence 10, Appl
2	21	47.7	9	3	US-08-339-339A-918 Sequence 28, Appl
3	21	47.7	9	5	PCT-US95-16415-25 Sequence 8, Appl
4	20	45.5	5	4	US-08-364-121-9 Sequence 9, Appl
5	20	45.5	5	4	US-08-496-692-9 Sequence 36, App
6	20	45.5	9	3	US-08-459-339A-336 Sequence 80, App
7	20	45.5	9	3	US-09-227-357-660 Sequence 8, Appl
8	19	43.2	8	3	US-03-121-964-5 Sequence 8, Appl
9	19	43.2	9	1	US-08-367-758B-8 Sequence 5, Appl
10	19	43.2	9	2	US-08-909-735-8 Sequence 11, Appl
11	18	40.9	6	4	US-09-179-558-52 Sequence 11, Appl
12	18	40.9	7	1	US-07-638-492-13 Sequence 4, Appl
13	18	40.9	7	3	US-08-556-418-14 Sequence 4, Appl
14	18	40.9	7	4	US-08-747-599A-4 Patent No. 5240848
15	18	40.9	7	6	5420848-3 Patent No. 5422249
16	18	40.9	9	1	US-08-338-634-12 Sequence 12, Appl
17	18	40.9	9	1	US-08-338-634-25 Sequence 30, Appl
18	18	40.9	9	1	US-08-338-634-30 Sequence 29, Appl
19	18	40.9	9	4	US-08-747-599A-29 Sequence 318, App
20	18	40.9	9	4	US-09-187-859-3138 Sequence 35, Appl
21	18	40.9	9	6	PCT-US95-16415-35 Patent No. 5200320
22	17	38.6	6	5	5200320-15 Sequence 8, Appl
23	17	38.6	7	1	US-08-180-209B-8 Sequence 8, Appl
24	17	38.6	7	1	US-08-385-745-8 Sequence 11, Appl
25	17	38.6	7	1	US-08-222-619-17 Sequence 26, Appl
26	17	38.6	7	4	US-08-747-599A-26 Sequence 26, Appl
27	17	38.6	7	4	US-08-747-599A-26 Sequence 26, Appl

28	17	38.6	7	4	US-08-485-388-8 Sequence 8, Appl
29	17	38.6	7	4	US-08-474-853-8 Sequence 8, Appl
30	17	38.6	7	4	US-09-166-205B-8 Sequence 8, Appl
31	17	38.6	7	5	PCT-US94-02629-8 Sequence 17, Appl
32	17	38.6	7	5	PCT-US95-04075-17 Sequence 28, Appl
33	17	38.6	9	1	US-08-338-634-28 Sequence 181, App
34	17	38.6	9	2	US-08-340-283-181 Sequence 320, App
35	17	38.6	9	2	US-08-146-028-320 Sequence 321, App
36	17	38.6	9	2	US-08-146-028-321 Sequence 326, App
37	17	38.6	9	2	US-08-146-028-322 Sequence 327, App
38	17	38.6	9	2	US-08-146-028-327 Sequence 328, App
39	17	38.6	9	2	US-08-146-028-328 Sequence 334, App
40	17	38.6	9	2	US-08-146-028-334 Sequence 330, App
41	17	38.6	9	4	US-08-723-425A-320 Sequence 321, App
42	17	38.6	9	4	US-08-723-425A-321 Sequence 326, App
43	17	38.6	9	4	US-08-723-425A-326 Sequence 327, App
44	17	38.6	9	4	US-08-723-425A-327 Sequence 328, App
45	17	38.6	9	4	US-08-723-425A-328 Sequence 328, App

ALIGNMENTS

RESULT 1
US-08-338-634-10
; Sequence 10, Application US/08338634
; Patent No. 5679641
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Peptides of human p53 protein for use
; TITLE OF INVENTION: in human T cell response inducing compositions, and
; TITLE OF INVENTION: human p53 protein-specific cytotoxic T-lymphocytes.
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann & Baron
; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11758
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/338, 634
; FILING DATE: 06-February-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL93/00102
; FILING DATE: 18-May-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-26
; TELEPHONE: (516) 822-3582
; TELEFAX: (516) 822-3550
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULAR TYPE: peptide
; HYPOTHEICAL: NO
; US-08-338-634-10
Query Match 47.7%, Score 21; DB 1; Length 9;
Best Local Similarity 37.5%; Fred. No. 1.9e+05;
Matches 3; Conservative 4; Mismatches 1; Indels 0;
Gy 1 ISEMPLQIY 8

Thu May 8 16:14:07 2003

us-09-658-621b-26.closed.ra1

Page 2

```
Db      2 LINKFCOL 9
      : : : : :
RESULT 2
US-08-159-339A-618
; Sequence 618, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Basteen
; TITLE OF INVENTION: HLA Binding peptides and their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: PASTESQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 618:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: anti acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-159-339A-618

Query Match      47.7%; Score 21; DB 3; Length 9;
Best Local Similarity 42.9%; Pred. No. 1.9e+05;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      3 EMTQIY 9
      | : : : |
      3 EFTVEM 9

Db      3 EFTVEM 9

RESULT 3
PCT-US95-16415-25
; Sequence 25, Application PC/TUS9516415
; GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; TITLE OF INVENTION: IN VIVO ACTIVATION OF TUMOR-SPECIFIC
; TITLE OF INVENTION: CYTOTOXIC T CELLS
; NUMBER OF SEQUENCES: 38
```

```
      ; CORRESPONDENCE ADDRESS:
      ; ADDRESSER: The Scripps Research Institute
      ; STREET: 10666 North Torrey Pines Road, TPC-8
      ; CITY: La Jolla
      ; STATE: California
      ; COUNTRY: US
      ; ZIP: 92037
      ; COMPUTER READABLE FORM:
      ; MEDIUM TYPE: Floppy disk
      ; COMPUTER: IBM PC compatible
      ; OPERATING SYSTEM: PC-DOS/MS-DOS
      ; SOFTWARE: Patent Release #1.0, Version #1.25
      ; CURRENT APPLICATION DATA:
      ; APPLICATION NUMBER: PCT/US95/16415
      ; FILING DATE: 13-DEC-1995
      ; CLASSIFICATION:
      ; PRIOR APPLICATION DATA:
      ; APPLICATION NUMBER: US 08/355,558
      ; FILING DATE: 14-DEC-1994
      ; ATTORNEY/AGENT INFORMATION:
      ; NAME: Logan, Rex H. C.
      ; REGISTRATION NUMBER: 13,950
      ; REFERENCE/DOCKET NUMBER: 433,1PC
      ; TELECOMMUNICATION INFORMATION:
      ; TELEPHONE: (619) 554-2937
      ; TELEFAX: (619) 554-6312
      ; INFORMATION FOR SEQ ID NO: 25:
      ; SEQUENCE CHARACTERISTICS:
      ; LENGTH: 9 amino acids
      ; TYPE: amino acid
      ; TOPOLOGY: linear
      ; MOLECULE TYPE: peptide
PCT-US95-16415-25

Query Match      47.7%; Score 21; DB 5; Length 9;
Best Local Similarity 37.5%; Pred. No. 1.9e+05;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 ISEMFQI 8
      : : : : :
      2 LINKFCOL 9

Db      2 LINKFCOL 9

RESULT 4
US-08-964-127-9
; Sequence 9, Application US/08964127
; Patent No. 6277565
; GENERAL INFORMATION:
; APPLICANT: Grandaearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: PASTESQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,127
; FILING DATE: 06-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crews, Ph.D., L. Lee
; REGISTRATION NUMBER: P-43,567
```

REFERENCE/DOCKET NUMBER: 07334/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-964-127-9

Query Match 45.5%; Score 20; DB 4; Length 5;
Best Local Similarity 80.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMF 5
||:|
Db 1 ISKMF 5

RESULT 5
US-09-496-692-9
Sequence 9, Application US/09496692
Patent No. 6313271
GENERAL INFORMATION:
APPLICANT: Grandearl, Andrew David John
TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/496,692
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/964,127
FILING DATE: 06-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Crews, Ph.D., L. Lee
REGISTRATION NUMBER: P-43,567
REFERENCE/DOCKET NUMBER: 07334/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-496-692-9

Query Match 45.5%; Score 20; DB 4; Length 5;
Best Local Similarity 80.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMF 5
||:|
Db 1 ISKMF 5

RESULT 6
US-08-159-339A-336
Sequence 336, Application US/08159339A
Patent No. 6037135
GENERAL INFORMATION:
APPLICANT: Kubo, Ralph T.
APPLICANT: Grey, Howard M.
APPLICANT: Sette, Alessandro
APPLICANT: Celis, Esteban
TITLE OF INVENTION: HLA Binding peptides and Their
TITLE OF INVENTION: Uses
NUMBER OF SEQUENCES: 1254
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,339A
FILING DATE: 29-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,666
FILING DATE: 07-AUG-1992
APPLICATION NUMBER: US 08/027,746
FILING DATE: 05-MAR-1993
APPLICATION NUMBER: US 08/103,396
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 018623-005030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
TELEX:
INFORMATION FOR SEQ ID NO: 336:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-159-339A-336

Query Match 45.5%; Score 20; DB 3; Length 9;
Best Local Similarity 22.2%; Pred. No. 1.9e+05;
Matches 2; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 ISEMFQIY 9
:::|:|
Db 1 LTKQYIMVY 9

RESULT 7
US-09-227-357-660
Sequence 660, Application US/09227357
Patent No. 6342581
GENERAL INFORMATION:
APPLICANT: Fischer et al.
TITLE OF INVENTION: 123 Human Secreted Proteins
FILE REFERENCE: P2010P1
CURRENT APPLICATION NUMBER: US/09/227,357
CURRENT FILING DATE: 1999-01-08
EARLIER APPLICATION NUMBER: PCT/US98/13684

EARLIER FILING DATE: 1998-07-07
 EARLIER APPLICATION NUMBER: 60/051,926
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,793
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,925
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,929
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,803
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,732
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,931
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,932
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,916
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,930
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,918
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,920
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,733
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,795
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,919
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,928
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/055,722
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,723
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,948
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,949
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,953
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,950
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,947
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,964
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/056,360
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,684
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,984
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,954
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/058,785
 EARLIER FILING DATE: 1997-09-12
 EARLIER APPLICATION NUMBER: 60/058,664
 EARLIER FILING DATE: 1997-09-12
 EARLIER APPLICATION NUMBER: 60/058,660
 EARLIER FILING DATE: 1997-09-12
 EARLIER APPLICATION NUMBER: 60/058,661
 EARLIER FILING DATE: 1997-09-12
 NUMBER OF SEQ ID NOS: 672
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 660
 LENGTH: 9
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-227-357-660

Query Match 45.5%; Score 20; DB 4; Length 9;
 Best Local Similarity 57.1%; Pred. No. 1.9e+05;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLO 7
 DB 1 IREIFLR 7

RESULT 8
 US-09-121-964-5
 Sequence 5; Application US/09121964
 Patent No. 6124447

GENERAL INFORMATION:
 APPLICANT: Natco, Shunji
 TITLE OF INVENTION: NOVEL ENZYME CATALYZING DEPHOSPHORYLATION
 FILE REFERENCE: 32290-144753
 CURRENT APPLICATION NUMBER: US/09/121,964
 CURRENT FILING DATE: 1998-07-24
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 5
 LENGTH: 8
 TYPE: PRT
 ORGANISM: Sarcophaga peregrina
 US-09-121-964-5

Query Match 43.2%; Score 19; DB 3; Length 8;
 Best Local Similarity 50.0%; Pred. No. 1.9e+05;
 Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 MELQIV 9
 DB 1 IFIOY 6

RESULT 9
 US-08-367-758B-8
 Sequence 8; Application US/08367758B
 Patent No. 5767235

GENERAL INFORMATION:
 APPLICANT: Muramatsu, Ryo
 APPLICANT: Sukesada, Akiko
 APPLICANT: Misawa, Satoru
 APPLICANT: Nakai, Eiko
 APPLICANT: Wada, Koichi
 APPLICANT: Nakano, Masaharu
 APPLICANT: Morikawa, Tadatoru
 TITLE OF INVENTION: Hybrid Variants, or Their Salts, Method
 TITLE OF INVENTION: For Their Production and Anticoagulants Having Said
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Birch, Stewart, Kolach & Birch, LLP
 STREET: P.O. Box 747
 CITY: Falls Church
 STATE: Virginia
 COUNTRY: USA
 ZIP: 22040-3487
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/367,758B
 FILING DATE: 03-JAN-1995
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy Jr., Gerald M.
 REGISTRATION NUMBER: 28,977
 REFERENCE/DOCKET NUMBER: 1516-188P


```
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..9
OTHER INFORMATION: /label= Example_3h_4h
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /product= "succinyl-glu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /product= "gln-oh"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /product= "tyr or tyr-SO3"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 7
OTHER INFORMATION: /product= "tyr or tyr-SO3"
US-08-367-758b-8

Query Match
43.2%; Score 19; DB 1; Length 9;
Best Local Similarity 57.1%; Pred. No. 1.9e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ISEMPLO 7
Db 3 IPEYVLQ 9

RESULT 10
US-08-909-735-8
Sequence 8, Application US/08909735
Patent No. 5880258
GENERAL INFORMATION:
APPLICANT: Muramatsu, Ryo
APPLICANT: Sukesada, Akiko
APPLICANT: Misawa, Satoru
APPLICANT: Nukui, Eriko
APPLICANT: Wada, Koichi
APPLICANT: Nakano, Masaharu
APPLICANT: Morikawa, Tadanori
TITLE OF INVENTION: Hindin Variants, or Their Salts, Method
TITLE OF INVENTION: For Their Production and Anticoagulants Having Said
TITLE OF INVENTION: Compounds as Active Ingredients
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSES: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/909,735
```

```
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/367,758
FILING DATE: 03-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1516-188P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..9
OTHER INFORMATION: /label= Example_3h_4h
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /product= "succinyl-glu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /product= "gln-oh"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /product= "tyr or tyr-SO3"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 7
OTHER INFORMATION: /product= "tyr or tyr-SO3"
US-08-909-735-8

Query Match
43.2%; Score 19; DB 2; Length 9;
Best Local Similarity 57.1%; Pred. No. 1.9e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ISEMPLO 7
Db 3 IPEYVLQ 9

RESULT 11
US-09-179-558-52
Sequence 52, Application US/09179558
Patent No. 6180612
GENERAL INFORMATION:
APPLICANT: Hockensmith, Joel W.
APPLICANT: Muthuswami, Rohini
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TITLE OF INVENTION: TARGETING DNA METABOLIC PROCESSES USING
TITLE OF INVENTION: AMINOGLYCOSIDE DERIVATIVES
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSES: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
```

SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/179,558
FILING DATE: 27-OCT-1998
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 09/060,470
FILING DATE: 15-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 60/063,898
FILING DATE: 31-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 9426-005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 7909090
TELEFAX: (212) 8699741
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-179-558-52

Query Match
Best Local Similarity 40.9%; Score 18; DB 4; Length 6;
Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 EMFL01 8
Db 1 ELYT01 6

RESULT 12
US-07-638-498-13
Sequence 13, Application US/07638492
GENERAL INFORMATION:
APPLICANT: Hodges, Robert S.
APPLICANT: Paranchych, William
APPLICANT: Irvin, Randall T.
APPLICANT: Lee, Kok K.
APPLICANT: Parimi, Saeed A.
APPLICANT: Zouman, Dick E.
APPLICANT: Doig, Peter C.
APPLICANT: Wong, Mah Y.
TITLE OF INVENTION: Pseudomonas Peptide Composition and
METHOD
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/638,492
FILING DATE: 04-JAN-1991
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 8900-0002.21

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: peptide 13
US-07-638-492-13

Query Match
Best Local Similarity 40.9%; Score 18; DB 1; Length 7;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 EMFL 6
Db 2 EMFL 5

RESULT 13
US-08-556-419-14
Sequence 14, Application US/08556419C
GENERAL INFORMATION:
APPLICANT: Ross, Christopher
APPLICANT: Li, Xiao-Jiang
APPLICANT: Li, Shi-Hua
APPLICANT: Sharp, Alan
APPLICANT: Latham, Anthony
APPLICANT: Wooley, Paul
APPLICANT: Snyder, Solomon
TITLE OF INVENTION: Huntingtin-associated protein
FILE REFERENCE: 01107,52271
CURRENT APPLICATION NUMBER: US/08/556,419C
CURRENT FILING DATE: 1995-11-09
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 7
TYPE: PPT
ORGANISM: Rattus norvegicus
US-08-556-419-14

Query Match
Best Local Similarity 40.9%; Score 18; DB 3; Length 7;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 LQIY 9
Db 4 LQIY 7

RESULT 14
US-08-747-599A-4
Sequence 4, Application US/08747599A
Patent No. 6214795
GENERAL INFORMATION:
APPLICANT: Benjamin, Howard et al.
TITLE OF INVENTION: Peptide Compounds Useful for Modulating
PCP Receptor Activity
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA

Thu May 8 16:14:07 2003

us-09-658-621b-26.closed.ra1

Page 7

ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,599A
FILING DATE: 12-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kara, Catherine J.
REGISTRATION NUMBER: 41,106
REFERENCE/DOCKET NUMBER: PPI-029
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: Internal
US-08-747-599A-4

Query Match 40.9%; Score 18; DB 4; Length 7;
Best Local Similarity 60.0%; Pred. No. 1.9e+05;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLDIV 9
Db 1 FLDIV 5

RESULT 15
5240848-3
Patent No. 5240848
APPLICANT: KECK, PAMELA J.; CONNOLLY, DANIEL T.; PEDER, JOSEPH
TITLE OF INVENTION: DNA SEQUENCES ENCODING HUMAN VASCULAR
PERMEABILITY FACTOR HAVING 189 AMINO ACIDS
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/337,037
FILING DATE: 10-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 274,061
FILING DATE: 21-NOV-1988
SEQ ID NO: 3:
LENGTH: 7
5240848-3

Query Match 40.9%; Score 18; DB 6; Length 7;
Best Local Similarity 40.0%; Pred. No. 1.9e+05;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLDIV 9
Db 1 FLDIV 5

Search completed: May 1, 2003, 23:24:59
Job time : 29 secs

Thu May 8 16:14:28 2003

us-09-658-621b-26.closed.rapb

Page 1

GenCore version 5.1.4 ps 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:24:07 ; Search time 46 Seconds
(without alignments)
16.882 Million cell updates/sec

Title: US-09-658-621B-26

Perfect score: 44

Sequence: 1 ISEMPIQIV 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 32825 seqs, 8628665 residues

Total number of hits satisfying chosen parameters: 30581

Minimum DB seq length: 0

Maximum DB seq length: 9

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.7	8	10	US-09-935-682-35	Sequence 35, Appl
2	47.7	9	9	US-09-277-074-25	Sequence 25, Appl
3	47.7	9	9	US-10-116-846B-35	Sequence 35, Appl
4	45.5	5	9	US-10-000-273-9	Sequence 9, Appl
5	45.5	6	9	US-09-814-604-45	Sequence 45, Appl
6	45.5	6	9	US-09-922-226-31	Sequence 31, Appl
7	45.5	6	9	US-09-922-226-184	Sequence 184, Appl
8	45.5	6	9	US-09-983-802-660	Sequence 660, Appl
9	45.5	9	10	US-09-753-831-14	Sequence 14, Appl
10	43.2	8	9	US-09-974-879-581	Sequence 581, Appl
11	43.2	9	9	US-09-809-638-34	Sequence 34, Appl
12	43.2	9	9	US-09-809-638-52	Sequence 52, Appl
13	43.2	9	9	US-09-809-638-135	Sequence 135, Appl
14	43.2	9	9	US-09-809-638-242	Sequence 242, Appl
15	43.2	9	9	US-09-803-638-452	Sequence 452, Appl
16	43.2	9	9	US-09-809-638-661	Sequence 661, Appl
17	43.2	9	9	US-10-116-846B-31	Sequence 31, Appl
18	43.2	9	10	US-09-780-053-33	Sequence 33, Appl
19	40.9	8	9	US-10-033-741-17	Sequence 17, Appl

20	18	40.9	8	9	US-10-140-164-10	Sequence 10, Appl
21	18	40.9	8	9	US-10-140-164-40	Sequence 40, Appl
22	18	40.9	9	1	US-08-821-739A-48	Sequence 48, Appl
23	18	40.9	9	9	US-09-826-290-24	Sequence 24, Appl
24	18	40.9	9	9	US-09-277-074-35	Sequence 35, Appl
25	18	40.9	9	9	US-09-791-389-98	Sequence 98, Appl
26	18	40.9	9	9	US-09-791-393-98	Sequence 98, Appl
27	18	40.9	9	10	US-09-897-107-21	Sequence 21, Appl
28	17	38.6	7	10	US-09-895-793-969	Sequence 969, Appl
29	17	38.6	7	10	US-09-822-827-969	Sequence 969, Appl
30	17	38.6	8	10	US-09-756-283A-41	Sequence 41, Appl
31	17	38.6	9	1	US-08-821-739A-50	Sequence 50, Appl
32	17	38.6	9	9	US-10-017-327-3	Sequence 3, Appl
33	17	38.6	9	9	US-10-017-327-5	Sequence 5, Appl
34	17	38.6	9	9	US-10-094-699-79	Sequence 79, Appl
35	17	38.6	9	9	US-10-094-699-80	Sequence 80, Appl
36	16	36.4	6	9	US-09-727-963A-38	Sequence 38, Appl
37	16	36.4	6	10	US-09-244-694-190	Sequence 190, Appl
38	16	36.4	7	9	US-10-124-880-39	Sequence 39, Appl
39	16	36.4	8	9	US-09-827-345-15	Sequence 15, Appl
40	16	36.4	8	9	US-10-062-710-137	Sequence 137, Appl
41	16	36.4	8	10	US-09-243-079-8	Sequence 8, Appl
42	16	36.4	8	10	US-09-243-079-9	Sequence 9, Appl
43	16	36.4	8	10	US-09-243-079-10	Sequence 10, Appl
44	16	36.4	8	10	US-09-243-079-11	Sequence 11, Appl
45	16	36.4	8	10	US-09-012-135A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-935-682-35
Sequence 35, Application US/09935682
Patent No. US20020059032A1
GENERAL INFORMATION:
APPLICANT: Societe de Conseils de Recherches et D'Applications Scientifiques
APPLICANT: Ferrer, Camara Y.
TITLE OF INVENTION: Rational Selection of Putative Peptides from Identified Nucleot.
FILE REFERENCE: 58767.000005
CURRENT APPLICATION NUMBER: US/09/935,682
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: 09/257,525
PRIOR FILING DATE: 1999-02-25
PRIOR APPLICATION NUMBER: PCT/FR00/00460
PRIOR FILING DATE: 2000-02-24
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35
LENGTH: 8
TYPE: PRT
ORGANISM: Homo sapiens
US-09-935-682-35
Query Match 47.7%; Score 21; DB 10; Length 8;
Best Local Similarity 80.0%; Pred. No. 2.9e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 ISEMP 5
DB 1 ISEMP 5
RESULT 2
US-09-277-074-25
Sequence 25, Application US/09277074
Publication No. US2003002820A1
GENERAL INFORMATION:
APPLICANT: Shennan, Linda A.
TITLE OF INVENTION: IN VIVO ACTIVATION OF TUMOR-SPECIFIC CYTOTOXIC T CELLS
FILE REFERENCE: SCR21555
CURRENT APPLICATION NUMBER: US/09/277,074

CURRENT FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 08/355,558
PRIOR FILING DATE: 1994-12-14
PRIOR APPLICATION NUMBER: PCT/US95/16415
PRIOR FILING DATE: 1995-12-14
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 25
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
US-09-277-074-25

Query Match 47.7%; Score 21; DB 9; Length 9;
Best Local Similarity 37.5%; Pred. No. 2.9e+05;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISMPLOI 8
Db 2 LNMFCOL 9

RESULT 3
US-10-116-846B-35
Sequence 35, Application US/10116846B
Publication No. US20030073102A1
GENERAL INFORMATION:
APPLICANT: KWOK, William
APPLICANT: GEHE, John
APPLICANT: REITONEN, Helena
APPLICANT: LIU, Andrew
TITLE OF INVENTION: METHODS OF MHC CLASS II EPTIPOPE MAPPING, DETECTION OF AUTOIMMUNE
TITLE OF INVENTION: AND ANTIGENS, AND AUTOIMMUNE TREATMENT
FILE REFERENCE: 20149-0005200S
CURRENT FILING DATE: 2002-09-16
PRIOR APPLICATION NUMBER: US 60/282,328
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: US 60/308,962
PRIOR FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn version 3.0
SEQ ID NO 35
LENGTH: 9
TYPE: PRT
ORGANISM: Herpes Simplex Virus 2
US-10-116-846B-35

Query Match 47.7%; Score 21; DB 9; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.9e+05;
Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 MFLQIY 9
Db 1 LFLMHLV 6

RESULT 4
US-10-000-273-9
Sequence 9, Application US/10000273
Patent No. US20020160386A1
GENERAL INFORMATION:
APPLICANT: Grandearl, Andrew David John
TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
MOLECULES
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA

ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/000,273
FILING DATE: 02-NO. US20020160386A1-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/964,127
FILING DATE: 06-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Crews, Ph.D., L. Lee
REGISTRATION NUMBER: P-43,567
REFERENCE/DOCKET NUMBER: 07334/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8908
TELEX: 200154
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-000-273-9

Query Match 45.5%; Score 20; DB 9; Length 5;
Best Local Similarity 80.0%; Pred. No. 2.9e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISMPF 5
Db 1 ISKMF 5

RESULT 5
US-09-814-604-45
Sequence 45, Application US/09814604
Publication No. US20030003517A1
GENERAL INFORMATION:
APPLICANT: Klein, Elliott S.
APPLICANT: Chandraratna Roshantha A.
TITLE OF INVENTION: Methods of Detecting Dissociated Nuclear
FILE REFERENCE: P-AR 4528
CURRENT APPLICATION NUMBER: US/09/814,604
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 45
LENGTH: 6
TYPE: PRT
ORGANISM: Homo sapiens
US-09-814-604-45

Query Match 45.5%; Score 20; DB 9; Length 6;
Best Local Similarity 33.3%; Pred. No. 2.9e+05;
Matches 2; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 4 MFLQIY 9
Db 1 LFLMHLV 6

RESULT 6
US-09-922-226-31
Sequence 31, Application US/09922226
Publication No. US2003007764A1
GENERAL INFORMATION:
APPLICANT: Zhao, Yi

```

; APPLICANT: Thacher, Scott M.
; APPLICANT: Xiao, Jia-Hao
; APPLICANT: Kusari, Jyotirmoy
; APPLICANT: Chandratna, Roshantha A.
; TITLE OF INVENTION: Methods of Screening for Compounds That
; FILE REFERENCE: P-AR 4681
; CURRENT APPLICATION NUMBER: US/09/922,226
; PRIOR FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: US 60/284,797
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-226-31
Query Match 45.5%; Score 20; DB 9; Length 6;
Best Local Similarity 33.3%; Pred. No. 2.9e+05;
Matches 2; Conservative 4; Mismatches 0; Gaps 0;

QY 4 MFLQIY 9
;|:|:|:
DB 1 LFLVLF 6

RESULT 7
US-09-922-226-184
; Sequence 184, Application US/09922226
; Publication No. US2003007664A1
; GENERAL INFORMATION:
; APPLICANT: Zhao, Yi
; APPLICANT: Thacher, Scott M.
; APPLICANT: Xiao, Jia-Hao
; APPLICANT: Kusari, Jyotirmoy
; APPLICANT: Chandratna, Roshantha A.
; TITLE OF INVENTION: Methods of Screening for Compounds That
; FILE REFERENCE: P-AR 4681
; CURRENT APPLICATION NUMBER: US/09/922,226
; PRIOR FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: US 60/284,797
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 184
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-226-184
Query Match 45.5%; Score 20; DB 9; Length 6;
Best Local Similarity 33.3%; Pred. No. 2.9e+05;
Matches 2; Conservative 4; Mismatches 0; Gaps 0;

QY 4 MFLQIY 9
;|:|:|:
DB 1 LFLVLF 6

RESULT 8
US-09-983-802-660
; Sequence 660, Application US/09983802
; Publication No. US20030022185A1
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/983,802
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357

; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 660
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-983-802-660

Query Match 45.5%; Score 20; DB 9; Length 9;
Best Local Similarity 57.1%; Pred. No. 2.9e+05;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ISEMFLQI 7
Db 1 IREIFLR 7

RESULT 9

US-09-753-831-14
Sequence 14, Application US/09753831
Patent No. US20020137683A1
GENERAL INFORMATION:
APPLICANT: Hogan, Kevin T.
TITLE OF INVENTION: C-SKI Oncogene-Derived Peptides for Prevention,
FILE REFERENCE: 26747-27
CURRENT APPLICATION NUMBER: US/09/753,831
PRIOR FILING DATE: 2001-01-03
PRIOR APPLICATION NUMBER: U.S. 60/174296
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: NO. US20020137683A1aPeptide
US-09-753-831-14

Query Match 45.5%; Score 20; DB 10; Length 9;
Best Local Similarity 33.3%; Pred. No. 2.9e+05;
Matches 3; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ISEMFLQI 9
Db 1 LERSRVRY 9

RESULT 10
US-09-974-879-581
Sequence 581, Application US/09974879
Publication No. US20030028003A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 125 Human Secreted Proteins
FILE REFERENCE: P2020P2
CURRENT APPLICATION NUMBER: US/09/974,879
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/239,893
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 09/818,683
PRIOR FILING DATE: 2001-03-26
PRIOR APPLICATION NUMBER: US 09/305,736
PRIOR FILING DATE: 1999-05-05
PRIOR APPLICATION NUMBER: PCT/US98/23435
PRIOR FILING DATE: 1998-11-04
PRIOR APPLICATION NUMBER: US 60/064,911
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,912
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,983
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,900
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,988
PRIOR FILING DATE: 1997-11-07

PRIOR APPLICATION NUMBER: US 60/064,987
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,908
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,984
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,985
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/066,094
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: US 60/066,100
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: US 60/066,089
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: US 60/066,095
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: US 60/066,090
NUMBER OF SEQ ID NOS: 611
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 581
LENGTH: 8
TYPE: PRT
ORGANISM: Homo sapiens
US-09-974-879-581

Query Match 43.2%; Score 19; DB 9; Length 8;
Best Local Similarity 42.9%; Pred. No. 2.9e+05;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 EMFLQI 9
Db 2 EIRGVF 8

RESULT 11
US-09-809-638-34
Sequence 34, Application US/09809638
Publication No. US2003005895A1
GENERAL INFORMATION:
APPLICANT: Mary Faris
APPLICANT: Pia M. Challita-Bid
APPLICANT: Steve Chappell Mitchell
APPLICANT: Daniel E.H. Afar
APPLICANT: Arthur B. Raitano
APPLICANT: Ava Jakobovits
TITLE OF INVENTION: 125P5C8: A TISSUE SPECIFIC PROTEIN
TITLE OF INVENTION: HIGHLY EXPRESSED IN VARIOUS CANCERS
FILE REFERENCE: 129.35US01
CURRENT APPLICATION NUMBER: US/09/809,638
PRIOR FILING DATE: 2001-03-14
NUMBER OF SEQ ID NOS: 746
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 34
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
US-09-809-638-34

Query Match 43.2%; Score 19; DB 9; Length 9;
Best Local Similarity 25.0%; Pred. No. 2.9e+05;
Matches 2; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SEMFLQI 9
Db 2 SEKYMLTF 9

RESULT 12
US-09-809-638-52
Sequence 52, Application US/09809638
Publication No. US2003005895A1
GENERAL INFORMATION:


```

; APPLICANT: Mary Paris
; APPLICANT: Pia M. Challita-Bid
; APPLICANT: Steve Chappell Mitchell
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Arthur B. Raitano
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: 125P5C8: A TISSUE SPECIFIC PROTEIN
; TITLE OF INVENTION: HIGHLY EXPRESSED IN VARIOUS CANCERS
; FILE REFERENCE: 129.35US01
; CURRENT APPLICATION NUMBER: US/09/809,638
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 746
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-809-638-52

```

```

Query Match          43.2%; Score 19; DB 9; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.9e+05;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ISEMFLQI 8
; : : : :
Db 2 LSPITLTI 9

```

```

RESULT 13
US-09-638-135
; Sequence 135, Application US/09809638
; Publication No. US20030059895A1
; GENERAL INFORMATION:
; APPLICANT: Mary Paris
; APPLICANT: Pia M. Challita-Bid
; APPLICANT: Steve Chappell Mitchell
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Arthur B. Raitano
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: 125P5C8: A TISSUE SPECIFIC PROTEIN
; TITLE OF INVENTION: HIGHLY EXPRESSED IN VARIOUS CANCERS
; FILE REFERENCE: 129.35US01
; CURRENT APPLICATION NUMBER: US/09/809,638
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 746
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 135
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-809-638-135

```

```

Query Match          43.2%; Score 19; DB 9; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.9e+05;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ISEMFLQI 8
; : : : :
Db 2 LSPITLTI 9

```

```

RESULT 14
US-09-809-638-242
; Sequence 242, Application US/09809638
; Publication No. US20030059895A1
; GENERAL INFORMATION:
; APPLICANT: Mary Paris
; APPLICANT: Pia M. Challita-Bid
; APPLICANT: Steve Chappell Mitchell
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Arthur B. Raitano
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: 125P5C8: A TISSUE SPECIFIC PROTEIN

```

```

; TITLE OF INVENTION: HIGHLY EXPRESSED IN VARIOUS CANCERS
; FILE REFERENCE: 129.35US01
; CURRENT APPLICATION NUMBER: US/09/809,638
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 746
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 242
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-809-638-242

```

```

Query Match          43.2%; Score 19; DB 9; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.9e+05;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ISEMFLQI 8
; : : : :
Db 2 LSPITLTI 9

```

```

RESULT 15
US-09-809-638-452
; Sequence 452, Application US/09809638
; Publication No. US20030059895A1
; GENERAL INFORMATION:
; APPLICANT: Mary Paris
; APPLICANT: Pia M. Challita-Bid
; APPLICANT: Steve Chappell Mitchell
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Arthur B. Raitano
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: 125P5C8: A TISSUE SPECIFIC PROTEIN
; TITLE OF INVENTION: HIGHLY EXPRESSED IN VARIOUS CANCERS
; FILE REFERENCE: 129.35US01
; CURRENT APPLICATION NUMBER: US/09/809,638
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 746
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 452
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-809-638-452

```

```

Query Match          43.2%; Score 19; DB 9; Length 9;
Best Local Similarity 25.0%; Pred. No. 2.9e+05;
Matches 2; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 SEMFLQIY 9
; : : : :
Db 2 SEKYWKLP 9

```

```

Search completed: May 1, 2003, 23:32:01
Job time : 47 secs

```


GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:10:31 ; Search time 4.96429 Seconds
(without alignments)
1168,869 Million cell updates/sec

Title: US-09-658-621b-26
Perfect score: 44
Sequence: 1 ISEWFLQIV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues
Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents_AA_Main:*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	9	20	US-09-658-621-26
2	44	100.0	10	20	US-09-658-621-29
3	44	100.0	292	26	US-10-221-279-12252
4	44	100.0	321	1	PCT-US00-05882-861
5	44	100.0	321	23	US-09-925-301-861
6	44	100.0	475	1	PCT-US01-30151-3

7	44	100.0	475	20	US-09-658-621-67	Sequence 67, Appl
8	44	100.0	495	20	US-09-658-621-2	Sequence 2, Appl
9	44	100.0	515	20	PCT-US02-07826-212	Sequence 212, App
10	44	100.0	515	24	US-10-097-340-212	Sequence 212, App
11	44	100.0	515	25	US-10-171-311-156	Sequence 156, App
12	44	100.0	1255	1	PCT-US02-19669-311	Sequence 311, Appl
13	44	100.0	1255	23	US-09-996-069-10	Sequence 10, Appl
14	44	100.0	1255	25	US-10-171-311-156	Sequence 158, App
15	44	100.0	1255	25	US-10-177-293-311	Sequence 311, Appl
16	41	93.2	508	17	US-09-366-670-20	Sequence 20, Appl
17	41	93.2	508	24	US-10-057-136-20	Sequence 20, Appl
18	36	81.8	448	21	US-09-760-446A-1839	Sequence 1839, Ap
19	36	81.8	448	26	US-10-206-664-1839	Sequence 1839, Ap
20	36	81.8	1083	21	US-09-791-537-101725	Sequence 101725,
21	36	81.8	1083	21	US-09-791-537-110555	Sequence 110555,
22	36	81.8	1083	21	US-09-791-537-115090	Sequence 115090,
23	36	81.8	1091	21	US-09-791-537-29746	Sequence 29746, A
24	36	81.8	1554	27	US-60-389-987-1395	Sequence 1395, Ap
25	36	81.8	1554	27	US-60-412-418-1395	Sequence 1395, Ap
26	34	77.3	545	23	US-09-966-728-2	Sequence 2, Appl
27	34	77.3	545	23	US-09-966-728-2	Sequence 2, Appl
28	33	75.0	57	19	US-09-513-996A-78504	Sequence 78504, A
29	33	75.0	61	19	US-09-513-996A-78166	Sequence 78166, A
30	33	75.0	71	19	US-09-513-996A-14787	Sequence 14787, A
31	33	75.0	71	20	US-09-620-394B-2999	Sequence 2999, Ap
32	33	75.0	71	23	US-09-935-625-24194	Sequence 24194, A
33	33	75.0	381	21	US-09-791-537-132877	Sequence 132877,
34	33	75.0	381	27	US-60-360-039-6373	Sequence 6373, Ap
35	33	75.0	839	21	US-09-708-427-16579	Sequence 16579, A
36	33	75.0	898	21	US-09-708-427-16579	Sequence 16579, A
37	33	75.0	955	21	US-09-708-427-16577	Sequence 16577, A
38	33	75.0	1005	19	US-09-513-996A-60874	Sequence 60874, A
39	33	75.0	1028	23	US-09-935-625-9787	Sequence 9787, Ap
40	33	75.0	1057	19	US-09-513-996A-60873	Sequence 60873, A
41	33	75.0	1080	23	US-09-935-625-9786	Sequence 9786, Ap
42	33	75.0	1114	19	US-09-513-996A-60872	Sequence 60872, A
43	32	72.7	1137	23	US-09-935-625-9785	Sequence 9785, A
44	32	72.7	97	21	US-09-758-441-244	Sequence 244, App
45	32	72.7	108	27	US-60-128-476-2574	Sequence 2574, Ap

ALIGNMENTS

RESULT 1
US-09-658-621-26
Sequence 26, Application US/09658621
GENERAL INFORMATION:
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Heukamp, Lukas Carl
APPLICANT: Offringa, Rienk
APPLICANT: Welliet, Cornelis Johanna Maria
APPLICANT: Acres, Bryce
APPLICANT: Thomas, Mireille
TITLE OF INVENTION: MGC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621
CURRENT FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 67
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 26
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621-26
Query Match 100.0%; Score 44; DB 20; Length 9;

Best Local Similarity 100.0%; Pred. No. 4.2e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
1 ISEMFLQIY 9

RESULT 2
US-09-658-621-29

Sequence 29, Application US/09658621
GENERAL INFORMATION:
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Heukamp, Lukas Carl
APPLICANT: Offringa, Rienk
APPLICANT: Melief, Cornelis Johanna Maria
APPLICANT: Acres, Bruce
APPLICANT: Thomas, Mireille
TITLE OF INVENTION: MUC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621
CURRENT FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 67
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 29
LENGTH: 10
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621-29

Query Match 100.0%; Score 44; DB 20; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.093;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
1 ISEMFLQIY 9

RESULT 3
US-10-221-279-12252

Sequence 12252, Application US/10221279
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 21272-046
CURRENT APPLICATION NUMBER: US/10/221,279
CURRENT FILING DATE: 2002-09-06
PRIOR APPLICATION NUMBER: 09/574,454
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: 09/519,705
PRIOR FILING DATE: 2000-03-07
NUMBER OF SEQ ID NOS: 12360
SOFTWARE: Custom
SEQ ID NO 12252
LENGTH: 292
TYPE: PRT
ORGANISM: Homo sapiens
US-10-221-279-12252

Query Match 100.0%; Score 44; DB 26; Length 292;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
110 ISEMFLQIY 118

RESULT 4
PCT-US00-05882-861

Sequence 861, Application PC/TUS0005882
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben,
TITLE OF INVENTION: Human Cancer Associated Gene Sequences and Polypeptides
FILE REFERENCE: P4106PCT
CURRENT APPLICATION NUMBER: PCT/US00/05882
CURRENT FILING DATE: 2000-03-08
EARLIER APPLICATION NUMBER: 60/124,270
EARLIER FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 861
LENGTH: 321
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (18)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US00-05882-861

Query Match 100.0%; Score 44; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
139 ISEMFLQIY 147

RESULT 5
US-09-925-301-861

Sequence 861, Application US/09925301
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P4106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 861
LENGTH: 321
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (18)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-861

Query Match 100.0%; Score 44; DB 23; Length 321;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
139 ISEMFLQIY 147

RESULT 6
PCT-US01-30151-3

Sequence 3, Application PC/TUS0130151
GENERAL INFORMATION:

APPLICANT: Genesance Pharmaceuticals, Inc.
APPLICANT: Anne, Chew
TITLE OF INVENTION: HAPLOYPES OF THE MUC1 GENE
FILE REFERENCE: MUC1 PNH-1408PCT
CURRENT APPLICATION NUMBER: PCT/US01/30151
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: 60/236,113
PRIOR FILING DATE: 2000-09-28
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 475
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-30151-3

Query Match 100.0%; Score 44; DB 1; Length 475;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLOIY 9
Db 293 ISEMFLOIY 301

RESULT 7
US-09-658-621-67
Sequence 67, Application US/09658621
GENERAL INFORMATION:
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Heukamp, Lukas Carl
APPLICANT: Offringa, Rienk
APPLICANT: Melief, Cornelis Johanna Maria
APPLICANT: Acres, Bruce
TITLE OF INVENTION: MUC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 67
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 67
LENGTH: 475
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621-67

Query Match 100.0%; Score 44; DB 20; Length 475;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLOIY 9
Db 293 ISEMFLOIY 301

RESULT 8
US-09-658-621-2
Sequence 2, Application US/09658621
GENERAL INFORMATION:
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Heukamp, Lukas Carl
APPLICANT: Offringa, Rienk
APPLICANT: Melief, Cornelis Johanna Maria
APPLICANT: Acres, Bruce
APPLICANT: Thomas, Mireille

TITLE OF INVENTION: MUC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 67
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 495
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621-2

Query Match 100.0%; Score 44; DB 20; Length 495;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLOIY 9
Db 313 ISEMFLOIY 321

RESULT 9
PCT-US02-07826-212
Sequence 212, Application PC/TUS0207826
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc. et al.
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
Title Of Invention: Assessment, Prevention, and Therapy of Ovarian Cancer
FILE REFERENCE: MRI-030PC
CURRENT APPLICATION NUMBER: PCT/US02/07826
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 212
LENGTH: 515
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-07826-212

Query Match 100.0%; Score 44; DB 1; Length 515;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLOIY 9
Db 333 ISEMFLOIY 341

RESULT 10
US-10-097-340-212
Sequence 212, Application US/10097340
GENERAL INFORMATION:
APPLICANT: John MONAHAN

```
APPLICANT: Manjula GANNAVARAPU
APPLICANT: Sebastian HOERSC
APPLICANT: Shubhangi KAMATKAR
APPLICANT: Steve G. KOVATS
APPLICANT: Rachel E. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEM
APPLICANT: Peter VEIBY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumei ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
FILE REFERENCE: MRI-030
CURRENT APPLICATION NUMBER: US/10/097,340
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 212
LENGTH: 515
TYPE: PRT
ORGANISM: Homo sapiens
US-10-097-340-212

Query Match 100.0%; Score 44; DB 24; Length 515;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLOIY 9
DB 333 ISEMFLOIY 341

RESULT 11
US-10-171-311-156
Sequence 156, Application US/1017311
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Chen, Yan
APPLICANT: Zhao, Xumei
APPLICANT: Mooshan, John
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Glatt, Karen
APPLICANT: Gannavarapu, Manjula
APPLICANT: Hoersb, Sebastian
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
FILE REFERENCE: MRI-035
CURRENT APPLICATION NUMBER: US/10/171,311
PRIOR FILING DATE: 2002-06-12
PRIOR APPLICATION NUMBER: US 60/298,159
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,155
PRIOR FILING DATE: 2001-06-13
```

```
PRIOR APPLICATION NUMBER: US 60/335,936
PRIOR FILING DATE: 2001-11-14
NUMBER OF SEQ ID NOS: 238
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 156
LENGTH: 515
TYPE: PRT
ORGANISM: Homo sapiens
US-10-171-311-156

Query Match 100.0%; Score 44; DB 25; Length 515;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLOIY 9
DB 333 ISEMFLOIY 341

RESULT 12
PCT-US02-19669-311
Sequence 311, Application PCT/US0219669
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc. et al.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038PC
CURRENT APPLICATION NUMBER: PCT/US02/19669
PRIOR FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 311
LENGTH: 1255
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-19669-311

Query Match 100.0%; Score 44; DB 1; Length 1255;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLOIY 9
DB 1073 ISEMFLOIY 1081

RESULT 13
US-09-996-069-10
Sequence 10, Application US/09996069
GENERAL INFORMATION:
APPLICANT: Bamdad, Cynthia
APPLICANT: Bamdad, R. Shoshana
TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBI
TITLE OF INVENTION: AND COMPOSITIONS AND METHODS FOR TREATMENT OF CANCER
FILE REFERENCE: MO1015/70071
CURRENT APPLICATION NUMBER: US/09/996,069
PRIOR FILING DATE: 2001-11-27
NUMBER OF SEQ ID NOS: 35
SOFTWARE: Patentin version 3.1
SEQ ID NO 10
LENGTH: 1255
```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-10

Query Match
Best Local Similarity 100.0%; Score 44; DB 23; Length 1255;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 1073 ISEMFLQTY 1081

RESULT 14
US-10-171-311-158
; Sequence 158, Application US/10171311
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatc, Karen
; APPLICANT: Ganavarapu, Manjula
; APPLICANT: Hoersch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 158
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-158

Query Match
Best Local Similarity 100.0%; Score 44; DB 25; Length 1255;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 1073 ISEMFLQTY 1081

RESULT 15
US-10-177-293-311
; Sequence 311, Application US/10177293
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatc, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Ganavarapu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Baet Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos

```

```

; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 311
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-311

Query Match
Best Local Similarity 100.0%; Score 44; DB 25; Length 1255;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 1073 ISEMFLQTY 1081

Search completed: May 1, 2003, 23:19:10
Job time : 5.96429 secs

```



```

FT NON TER 5 5 72CB19C9C0300000 CRC64;
SQ SEQUENCE 5 AA; 654 MW;

Query Match
Best Local Similarity 27.3%; Score 12; DB 1; Length 5;
Matches 1; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 MFLQI 8
|:|:
|:|:
DB 1 IFFEV 5

RESULT 7
CIP2_MYTED STANDARD; PRT; 6 AA.
AC P13737;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Contraction-inhibiting peptide II (MIP II).
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Periomorphia; Mytiloida;
OC Mytiloidae; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE.
RC TISSUE=Pedal ganglion;
RX MEDLINE=88240357; PubMed=3377776;
RA Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT "Structures and actions of Mytilus inhibitory peptides.";
RL Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC -1- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN MUSCLES.
DR -1- SIMILARITY: TO MIP I.
DR PIR; B27696; B27696.
KW Hormone; Amidation.
FT MOD_RES 6
SQ SEQUENCE 6 AA; 621 MW; 72C9C6876DD81000 CRC64;

Query Match
Best Local Similarity 27.3%; Score 12; DB 1; Length 6;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 MFL 6
|:|:
|:|:
DB 4 MEV 6

RESULT 8
E101_LITRU STANDARD; PRT; 6 AA.
ID E101_LITRU
AC P82096;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Electrin 1.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hyllidae;
OC Pelodyadinae; Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE.
RC TISSUE=Skin secretion;
RX Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT "Peptides from the skin glands of the Australian buzzing tree frog Litoria electrica. Comparison with the skin peptides from Litoria rubella";
RT Aust. J. Chem. 52:639-645(1999).
CC -1- SUBCELLULAR LOCATION: SECRETED.
KW Amphibian skin; Amidation.
FT MOD_RES 6
SQ SEQUENCE 6 AA; 792 MW; 6683704772C9A000 CRC64;

```

```

Query Match
Best Local Similarity 27.3%; Score 12; DB 1; Length 6;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 FFLQIY 9
|:|:
|:|:
DB 1 FVPIW 5

RESULT 9
ALU6_CYDPO STANDARD; PRT; 8 AA.
ID ALU6_CYDPO
AC P82157;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cydia pomonella 6.
OS Cydia pomonella (Coddling moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptysa; Tortricidae; Tortricidae; Olethreutinae; Cydia.
OX NCBI_TaxID=82600;
RN [1]
RP SEQUENCE.
RC TISSUE=larva;
RX MEDLINE=98054539; PubMed=9392829;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA Davey M., East P.D., Thorpe A.;
RT "Lepidopteran peptides of the allostatin superfamily.";
RL Peptides 18:1301-1309(1997).
CC -1- SIMILARITY: BELONGS TO THE ALLOSTATIN FAMILY.
KW Neuropeptide; Amidation.
FT MOD_RES 8
SQ SEQUENCE 8 AA; 936 MW; 0B2879C458573767 CRC64;

Query Match
Best Local Similarity 27.3%; Score 12; DB 1; Length 8;
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 6 LQIY 9
|:|:
|:|:
DB 1 LPIY 4

RESULT 10
FAR2_CALVO STANDARD; PRT; 9 AA.
ID FAR2_CALVO
AC P41857;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Calliphras vombitoria 2.
OS Calliphora vomitoria (Blue blowfly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Oestroidea; Calliphoridae; Calliphora.
OX NCBI_TaxID=27454;
RN [1]
RP SEQUENCE.
RC TISSUE=Thoracic ganglion;
RX MEDLINE=92196111; PubMed=1549595;
RA Duvé H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA Rehfeld J.F., Thorpe A.;
RT "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2 neuropeptides (designated calliphrasins) from the blowfly Calliphora vomitoria.";
RT Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC -1- FUNCTION: ABLE TO INDUCE FLUID SECRETION FROM THE ISOLATED SALIVARY GLAND OF CALLIPHORA.
CC -1- SIMILARITY: BELONGS TO THE FARP (FARFAMIDE RELATED PEPTIDE) FAMILY.
DR PIR; B41978; B41978.

```

```

KW Neuropeptide; Amidation.
FT MOD RES 9 AA; 1128 MW; 29D00699CAB6C5A7 CRC64;
SQ SEQUENCE 2

Query Match
Best Local Similarity 33.3%; Score 12; DB 1; Length 9;
Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 SEMFLO 7
| | |
| | |
DB 3 SODPMR 8

RESULT 11
FAR3 CALVO STANDARD; PRT; 9 AA.
AC P41658;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Callifamamide 3.
OS Calliphora vomitoria (Blue blowfly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Oestroidea; Calliphoridae; Calliphora.
OX NCBI_TaxId=27454;
RN [1]
RP SEQUENCE.
RC TISSUE=Thoracic ganglion;
RX MEDLINE=9219611; PubMed=1549595;
RA Duvé H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA Rehfeld J.F., Thorpe A.;
RT "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT neuropeptides (designated callifamamides) from the blowfly
RT Calliphora vomitoria."
RL Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC -1- FUNCTION: ABLE TO INDUCE FLUID SECRETION FROM THE ISOLATED
CC SALIVARY GLAND OF CALLIPHORA.
CC -1- SIMILARITY: BELONGS TO THE FAR3 (FAMFAMIDE RELATED PEPTIDE)
CC FAMILY.
CC PIR; C41978; C41978.
KW Neuropeptide; Amidation.
FT MOD RES 9 AA; 1114 MW; 2F0B0699CAB6C5A7 CRC64;
SQ SEQUENCE 9 AA; 1114 MW; 2F0B0699CAB6C5A7 CRC64;

Query Match
Best Local Similarity 33.3%; Score 12; DB 1; Length 9;
Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 SEMFLO 7
| | |
| | |
DB 3 SODPMR 8

RESULT 12
FLRF HIRME STANDARD; PRT; 4 AA.
AC P42561;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE FLRFamide.
OS Hirudo medicinalis (Medicinal leech), and
OS Helisoma trivolvis (Snail).
OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinea; Hirudinea;
OC Arynchobdellida; Hirudinoformes; Hirudindae; Hirudo.
OX NCBI_TaxId=6421, 27815;
RN [1]
RP SEQUENCE.
RC SPECIES=H. medicinalis;
RX MEDLINE=9215954; PubMed=1686933;
RA Evans 3.D., Pohl J., Kartsolis M.A., Calabrese R.L.;
RT "Identification of Rfamide neuropeptides in the medicinal leech."

```

```

RL Peptides 12:897-908(1991).
RN [2]
RP SEQUENCE.
RC SPECIES=H.trivolvis; TISSUE=Kidney;
RX MEDLINE=9428647; PubMed=7912428;
RA Madrid K.P., Price D.A., Greenberg M.J., Khan H.R., Saleuddin A.S.M.;
RT "FMRamide-related peptides from the kidney of the snail, Helisoma
RT trivolvis."
RL Peptides 15:31-36(1994).
CC -1- SIMILARITY: BELONGS TO THE FAR3 (FMRFAMIDE RELATED PEPTIDE)
CC FAMILY.
KW Neuropeptide; Amidation.
FT MOD RES 4 AA; 582 MW; 69D40729A0000000 CRC64;
SQ SEQUENCE 4 AA; 582 MW; 69D40729A0000000 CRC64;

Query Match
Best Local Similarity 66.7%; Score 11; DB 1; Length 4;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 FLO 7
| | |
| | |
DB 1 FLR 3

RESULT 13
FLRN ANTFL STANDARD; PRT; 4 AA.
AC P58707;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Antho-RNamide.
OS Anthopleura elegantissima (Sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;
OC Nynanthaea; Actiniidae; Anthopleura.
OX NCBI_TaxId=6110;
RN [1]
RP SEQUENCE, AND MASS-SPECTROMETRY.
RX MEDLINE=90319122; PubMed=1973541;
RA Grimmelikhuijzen C.J.P., Rinehart K.L. Jr., Jacob E., Graff D.,
RA Reincheid R.K., Norhacker H.-P., Staley A.L.;
RT "Isolation of L-3-phenylacetyl-L-Asn-Arg-Asn-NH2 (Antho-RNamide), a sea
RT anemone neuropeptide containing an unusual amino-terminal blocking
RT group."
RL Proc. Natl. Acad. Sci. U.S.A. 87:5410-5414(1990).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Neuron-specific.
CC -1- MASS SPECTROMETRY: MW=549.3; METHOD=FAE.
KW Neuropeptide; Amidation.
FT MOD RES 1 AA; 549 MW; 64540739A0000000 CRC64;
FT MOD RES 4 AA; 549 MW; 64540739A0000000 CRC64;
SQ SEQUENCE 4 AA; 549 MW; 64540739A0000000 CRC64;

Query Match
Best Local Similarity 25.0%; Score 11; DB 1; Length 4;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 FLO 7
| | |
| | |
DB 1 FLR 3

RESULT 14
PSK DAUCA STANDARD; PRT; 5 AA.
AC P58261;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phytosulfokine-alpha (PSK-alpha) [Contains: Phytosulfokine-beta (PSK-
DE beta)].
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```


CURRENT APPLICATION NUMBER: US/09/658,621B
CURRENT FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn version 3.1
SEQ ID NO 29
LENGTH: 10
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621B-29

Query Match 100.0%; Score 44; DB 5; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | | | | | | |
| | | | | | | | | |
Db 82 ISEMFLQIY 90

RESULT 3
US-60-452-680-12836
Sequence 12836, Application US/60452680
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
APPLICANT: GRUPE, Andrew
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001450
CURRENT APPLICATION NUMBER: US/60/452,680
CURRENT FILING DATE: 2003-03-07
NUMBER OF SEQ ID NOS: 116213
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12836
LENGTH: 264
TYPE: PRT
ORGANISM: Homo sapiens
US-60-452-680-12836

Query Match 100.0%; Score 44; DB 7; Length 264;
Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | | | | | | |
| | | | | | | | | |
Db 82 ISEMFLQIY 90

RESULT 4
US-60-453-135-8037
Sequence 8037, Application US/60453135
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
APPLICANT: IAKOBOVA, Olga
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001456
CURRENT APPLICATION NUMBER: US/60/453,135
CURRENT FILING DATE: 2003-03-10
NUMBER OF SEQ ID NOS: 82762
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8037
LENGTH: 264
TYPE: PRT
ORGANISM: Homo sapiens
US-60-453-135-8037

Query Match 100.0%; Score 44; DB 7; Length 264;

Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | | | | | | |
| | | | | | | | | |
Db 82 ISEMFLQIY 90

RESULT 5
US-60-453-050-8037
Sequence 8037, Application US/60453050
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
APPLICANT: LUXE, May
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001457
CURRENT APPLICATION NUMBER: US/60/453,050
CURRENT FILING DATE: 2003-03-10
NUMBER OF SEQ ID NOS: 82762
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8037
LENGTH: 264
TYPE: PRT
ORGANISM: Homo sapiens
US-60-453-050-8037

Query Match 100.0%; Score 44; DB 7; Length 264;
Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | | | | | | |
| | | | | | | | | |
Db 82 ISEMFLQIY 90

RESULT 6
US-09-658-621B-67
Sequence 67, Application US/09658621B
GENERAL INFORMATION:
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Henkamp, Lukas Carl
APPLICANT: Offridga, Rienk
APPLICANT: Melief, Cornelis Johanna Maria
APPLICANT: Acres, Bruce
TITLE OF INVENTION: WGC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621B
CURRENT FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn version 3.1
SEQ ID NO 67
LENGTH: 475
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621B-67

Query Match 100.0%; Score 44; DB 5; Length 475;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | | | | | | |
| | | | | | | | | |
Db 293 ISEMFLQIY 301

```
RESULT 7
US-10-348-119-309
; Sequence 309, Application US/10348119
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 NP
; CURRENT APPLICATION NUMBER: US/10/348,119
; CURRENT FILING DATE: 2003-01-17
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 309
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-348-119-309

Query Match
Best Local Similarity 100.0%; Score 44; DB 6; Length 475;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 293 ISEMFLQTY 301

RESULT 8
US-60-422-176-59
; Sequence 59, Application US/60422176
; GENERAL INFORMATION:
; APPLICANT: Pharmacia Corporation
; TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED GENES INVOLVED IN CANCER, THE POLYPEPTIDE
; TITLE OF INVENTION: ENCODED THEREBY, AND METHODS OF USING THE SAME
; FILE REFERENCE: 01040 PR
; CURRENT APPLICATION NUMBER: US/60/422,176
; CURRENT FILING DATE: 2002-10-29
; NUMBER OF SEQ ID NOS: 3114
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-422-176-59

Query Match
Best Local Similarity 100.0%; Score 44; DB 7; Length 475;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 293 ISEMFLQTY 301

RESULT 9
US-60-452-680-12835
; Sequence 12835, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12835
; LENGTH: 475
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-12835

Query Match
Best Local Similarity 100.0%; Score 44; DB 7; Length 475;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 293 ISEMFLQTY 301

RESULT 10
US-60-453-135-8036
; Sequence 8036, Application US/60453135
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001456
; CURRENT APPLICATION NUMBER: US/60/453,135
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8036
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-135-8036

Query Match
Best Local Similarity 100.0%; Score 44; DB 7; Length 475;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 293 ISEMFLQTY 301

RESULT 11
US-60-453-050-8036
; Sequence 8036, Application US/60453050
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001457
; CURRENT APPLICATION NUMBER: US/60/453,050
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8036
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-050-8036

Query Match
Best Local Similarity 100.0%; Score 44; DB 7; Length 475;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 293 ISEMFLQTY 301

RESULT 12
US-09-658-621b-2
; Sequence 2, Application US/09658621B
; GENERAL INFORMATION:
```

```
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Henkamp, Lukas Carl
APPLICANT: Offringa, Rijk
APPLICANT: Mellet, Cornelis Johanna Maria
APPLICANT: Acres, Bruce
APPLICANT: Thomas, Mireille
TITLE OF INVENTION: MUC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621B
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 495
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621B-2
```

```
Query Match          100.0%; Score 44; DB 5; Length 495;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 ISEMFLQIY 9
Db 313 ISEMFLQIY 321
```

```
RESULT 13
PCT-US02-19669A-311
Sequence 311, Application PC/TUS0219669A
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc. et al.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
FILE REFERENCE: MRI-038PC
CURRENT APPLICATION NUMBER: PCT/US02/19669A
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/239,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 311
LENGTH: 1255
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-19669A-311
```

```
Query Match          100.0%; Score 44; DB 1; Length 1255;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 ISEMFLQIY 9
Db 1073 ISEMFLQIY 1081
```

RESULT 14

```
PCT-IL02-00255-2
Sequence 2, Application PC/IL0200255
GENERAL INFORMATION:
APPLICANT: Ramot University Authority for Applied Research
APPLICANT: WRESCHNER, Daniel
APPLICANT: YOSHIKAWA, Naoyuki
TITLE OF INVENTION: Peptides and Antibodies to MUC 1 Proteins
FILE REFERENCE: P-3891-PC
CURRENT APPLICATION NUMBER: PCT/IL02/00255
PRIOR FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: 60/279,408
PRIOR FILING DATE: 2001-03-29
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 1255
TYPE: PRT
ORGANISM: Homo sapiens
PCT-IL02-00255-2
```

```
Query Match          100.0%; Score 44; DB 1; Length 1255;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 ISEMFLQIY 9
Db 1073 ISEMFLQIY 1081
```

```
RESULT 15
US-09-949-016-9618
Sequence 9618, Application US/09949016
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CD001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9618
LENGTH: 1041
TYPE: PRT
ORGANISM: Human
US-09-949-016-9618
```

```
Query Match          81.8%; Score 36; DB 5; Length 1041;
Best Local Similarity 77.8%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 ISEMFLQIY 9
Db 679 ISEMFLQIY 687
```

Search completed: May 1, 2003, 23:21:01
Job time : 1.91071 secs

GenCore version 5.1.4.P5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:14:31 / Search time 28 Seconds
(without alignments)
66.229 Million cell updates/sec

Title: US-09-658-621b-26

Perfect score: 44

Sequence: 1 ISEWFLQIV 9

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 686

Minimum DB seq length: 0

Maximum DB seq length: 9

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SPTREMBL 21:*

1: sp. archaea:*

2: sp. bacteria:*

3: sp. fungi:*

4: sp. human:*

5: sp. invertebrate:*

6: sp. mammal:*

7: sp. mhc:*

8: sp. organelle:*

9: sp. phage:*

10: sp. plant:*

11: sp. rodent:*

12: sp. virus:*

13: sp. vertebrate:*

14: sp. unclassified:*

15: sp. virus:*

16: sp. bacteriophage:*

17: sp. archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	40.9	9	2	P72345
2	16	36.4	8	2	O9RQ49
3	15	34.1	8	7	O84V76
4	15	34.1	9	7	O78226
5	15	34.1	9	8	O94X56
6	15	34.1	9	8	O94X56
7	15	34.1	9	8	O94NB1
8	15	34.1	9	8	O94NB1
9	15	34.1	9	8	O94NB1
10	15	34.1	9	8	O94NB1
11	13	29.5	8	2	O68485
12	13	29.5	8	2	P83158
13	13	29.5	8	5	O9VRD2
14	13	29.5	8	15	O98YK9
15	13	29.5	9	2	O93738
16	13	29.5	9	2	O93738

17	13	29.5	9	4	P78484	P78484 homo sapien
18	13	29.5	9	4	O14277	O14277 homo sapien
19	13	29.5	9	11	O62530	O62530 mus spretus
20	13	29.5	9	12	O88953	O88953 vaccinia vi
21	12	27.3	8	2	O45889	O45889 clostridium
22	12	27.3	8	2	O938P2	O938P2 pseudomonas
23	12	27.3	8	2	O37854	O37854 bacteriophage
24	12	27.3	8	10	O40530	O40530 hiccotiana t
25	12	27.3	9	2	O47556	O47556 escherichia
26	12	27.3	9	2	P83157	P83157 anabaena sp
27	12	27.3	9	4	O9UKJ6	O9UKJ6 homo sapien
28	12	27.3	9	4	O9UC36	O9UC36 homo sapien
29	12	27.3	9	11	O9QWC2	O9QWC2 mus musculu
30	12	27.3	9	11	O92012	O92012 mus musculu
31	12	27.3	9	12	O83622	O83622 murray vail
32	11	25.0	7	2	O54248	O54248 streptomyce
33	11	25.0	7	5	P83274	P83274 macrobrachi
34	11	25.0	8	2	O9RQ57	O9RQ57 buchnera ap
35	11	25.0	8	2	O96429	O96429 thermus the
36	11	25.0	8	2	P72279	P72279 rhodococcus
37	11	25.0	8	4	O15890	O15890 homo sapien
38	11	25.0	8	4	O16428	O16428 homo sapien
39	11	25.0	8	5	O9UB13	O9UB13 albinaria h
40	11	25.0	8	5	P83275	P83275 macrobrachi
41	11	25.0	8	5	P83316	P83316 penaeus mon
42	11	25.0	8	5	O9MYL5	O9MYL5 pongo pygma
43	11	25.0	8	6	O9XNP8	O9XNP8 boophilus m
44	11	25.0	8	8	O9GD47	O9GD47 hydraestele
45	11	25.0	8	8	O9GC24	O9GC24 nenga puni1

ALIGNMENTS

RESULT 1	P72345	PRELIMINARY;	PRT;	9 AA.
AC	P72345:			
DT	01-FEB-1997 (TREMBLrel. 02, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)			
DE	Hypothetical 1.0 kDa protein (fragment).			
GN	TABA.			
OS	Pseudomonas syringae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OX	Pseudomonas.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-93167809; PubMed-7679566;			
RA	Barta T.M., Kinscherf T.G., Uchytel T.F., Willis D.K.;			
RT	"DNA sequence and transcriptional analysis of the tda gene required			
RT	for tabtoxin biosynthesis by Pseudomonas syringae."			
RL	Appl. Environ. Microbiol. 59:458-466(1993).			
DR	EMBL; S54909; AAB25381.2; -			
KW	Hypothetical protein.			
FT	NON TER			
SQ	SEQUENCE 9 AA; 1037 MW; 283409D5B805B047 CRC64;			
Query Match	40.9%;	Score 18;	DB 2;	Length 9;
Best Local Similarity	80.0%;	Pred. No. 6.7e+05;		
Matches 4;	Conservative	0;	Mismatches 1;	Indels 0;
QY	1 ISEWF 5			
DB	3 ISEWF 7			
RESULT 2	O9RQ49	PRELIMINARY;	PRT;	8 AA.
ID	O9RQ49:			
AC	O9RQ49:			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE NIFS protein homolog (Fragment).
 GN NIFS.
 OS Buchnera aphidicola.
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 CX NCBI_TaxId=9;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20022990; PubMed=10555290;
 RA Clark M.A., Moran N.A., Baumann P.;
 RT "Sequence evolution in bacterial endosymbionts having extreme base
 composition.";
 RL Mol. Biol. Evol. 16:1586-1598(1999).
 DR EMBL; AF130814; AAF13805.1; -.
 FT NON TER 8
 SQ SEQUENCE 8 AA; 992 MW; F3A73B5047672336 CRC64;

Query Match 36.4%; Score 16; DB 2; Length 8;
 Best Local Similarity 66.7%; Pred. No. 6.7e+05;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 MFLQIY 9
 DB 1 MKLPIY 6

RESULT 3
 Q94VF6 PRELIMINARY; PRT; 8 AA.

DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DE Cytochrome c oxidase subunit I (Fragment).
 GN COI.
 OS Varanus jobiensis.
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Anguimorpha; Varanidae; Varanus.
 CX NCBI_TaxId=169843;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Ast J.C.;
 RT "Mitochondrial DNA evidence and evolution in Varanoidea (Squamata).";
 RL Cladistics 17:0-0(2001).
 DR EMBL; AF407507; AAL10075.1; -.
 KW Mitochondrion.
 FT NON TER 8
 SQ SEQUENCE 8 AA; 1144 MW; EFD729DB436411A6 CRC64;

Query Match 34.1%; Score 15; DB 8; Length 8;
 Best Local Similarity 100.0%; Pred. No. 6.7e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 FLQ 7
 DB 6 FLQ 8

RESULT 4
 Q78226 PRELIMINARY; PRT; 9 AA.

DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DE Lymphocyte antigen (Fragment).
 GN Mm musculus (Mouse).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxId=10090;
 RN (1)

RP SEQUENCE FROM N.A.
 RC STRAIN=CRO437; TISSUE=KIDNEY;
 RX MEDLINE=88084418; PubMed=3692165;
 RA Golubic M., Budimir O., Schoepfer R., Kasahara M., Mayer W.E.,
 RA Figueroa F., Klein J.;
 RT "Nucleotide sequence analysis of class II genes borne by mouse t
 RT chromosomes.";
 RL Genet. Res. 50:137-146(1987).
 DR EMBL; J38590; AAB57294.1; -.
 FT NON TER 1
 SQ SEQUENCE 9 AA; 1176 MW; 86CB1412C729C33A CRC64;

Query Match 34.1%; Score 15; DB 7; Length 9;
 Best Local Similarity 60.0%; Pred. No. 6.7e+05;
 Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SEMFL 6
 DB 1 SDRFL 5

RESULT 5
 Q94XE6 PRELIMINARY; PRT; 9 AA.

DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DE Cytochrome c oxidase subunit III (Fragment).
 GN COX3.
 OS Tectocoris diophthalmus (cotton harlequin bug).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Paraneoptera; Hemiptera; Euhemiptera;
 OC Pentatomera; Pentheteroptera; Pentatomomorpha; Pentatomidae;
 OC Pentatomidae; Tectocoris.
 CX NCBI_TaxId=159956;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21396409; PubMed=11504862;
 RT "Increased rate of gene rearrangement in the mitochondrial genomes of
 RT three orders of hemipteroid insects.";
 RL Mol. Biol. Evol. 18:1828-1832(2001).
 DR EMBL; AF335990; AAK55283.1; -.
 KW Mitochondrion.
 FT NON TER 1
 SQ SEQUENCE 9 AA; 1206 MW; A2C563636B5041A6 CRC64;

Query Match 34.1%; Score 15; DB 8; Length 9;
 Best Local Similarity 40.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLQIY 9
 DB 1 YMTIY 5

RESULT 6
 Q94NB2 PRELIMINARY; PRT; 9 AA.

DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DE Cytochrome oxidase subunit III (Fragment).
 GN COXII.
 OS Microtus murinus (lesser mouse lemur).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Strepsirhini; Cheirogaleidae;
 OC Microtus.

QY 1 ISEMFLQIY 9
 Db 729 ISEMFLQIY 737

RESULT 5

Probable Arac-type regulator [imported] - Sinorhizobium meliloti (strain 1021) megaplast
 C:Species: Sinorhizobium meliloti
 C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: D95283
 R:Barner, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bove
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
 A>Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
 A:Reference number: A95262; MUID:21396509; PMID:11461432
 A:Accession: D95283
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-333 <RUR>
 A:Cross-references: GB:AE006469; PIDN:AAK64830.1; PID:G14523242; GSPDB:GN00165
 A:Experimental source: strain 1021, megaplastid pSymA
 R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 J.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kles, E.; Komp, C.; LaLaure,
 hebaul, P.; Vandenbol, M.; Vortholter, F.J.; Weidner, S.; Weller, D.H.; Wong, K.; Yeh, K.
 A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: Sma0319
 A:Genome: plasmid

Query Match 75.0%; Score 33; DB 2; Length 333;
 Best Local Similarity 62.5%; Pred. No. 18;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
 Db 65 SDMPVQY 72

RESULT 6

T29826
 Hypothetical protein C5SH.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T29826
 R:Pauley, A.; Gairing, S.
 submitted to the EMBL Data Library, April 1996
 A:Description: The sequence of C. elegans comid C5SH.
 A:Reference number: Z20694
 A:Accession: T29826
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-381 <PAU>
 A:Cross-references: EMBL:U55567; PIDN:AAA97988.1; GSPDB:GN00023; CESP:C5SH.2
 A:Experimental source: strain Bristol N2; clone C5SH1
 C:Genetics:
 A:Gene: C5SH.2
 A:Map position: 5
 A:Intons: 21/2; 57/2; 103/3; 156/2; 182/2; 238/3; 268/3; 320/1; 347/2
 C:Superfamily: GTP-binding regulatory protein Gs alpha chain

Query Match 75.0%; Score 33; DB 2; Length 381;
 Best Local Similarity 66.7%; Pred. No. 21;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
 Db 169 MSEMFLKIF 177

RESULT 7
 A47447
 calcium channel protein alpha-1 chain (variant doe-1) - electric ray (Discopyge ommata)

C:Species: Discopyge ommata
 C>Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
 C:Accession: A47447
 R:Horne, W.A.; Ellnor, P.T.; Inman, I.; Zhou, M.; Tsiem, R.W.; Schwarz, T.L.
 Proc. Natl. Acad. Sci. U.S.A. 90, 3787-3791, 1993
 A>Title: Molecular diversity of Ca(2+) channel alpha 1 subunits from the marine ray Disc
 A:Reference number: A47447; MUID:93248175; PMID:7683405
 A:Accession: A47447
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-2223 <HOR>
 A>Note: sequence extracted from NCBI backbone (NCBIP:130671)
 C:Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match 75.0%; Score 33; DB 2; Length 2223;
 Best Local Similarity 75.0%; Pred. No. 1.4e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
 Db 535 MSEMFLKIF 542

RESULT 8

T13980
 calcium channel alpha-1 chain - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
 C:Accession: T13980
 R:Smith, L.A.; Wang, X.J.; Pelicci, A.P.; Neumann, E.K.; Hall, L.M.; Hall, J.C.
 J. Neurosci. 16, 7868-7879, 1996
 A>Title: A Drosophila calcium channel alpha-1 subunit gene maps to a genetic locus ass
 A:Reference number: Z17844; MUID:9741514; PMID:8987815
 A:Accession: T13980
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1851 <SMI>
 A:Cross-references: EMBL:U55776; NID:G1737063; PID:G1737064; PIDN:AA47406.1
 C:Genetics:
 A:Gene: calt
 A:Cross-references: FlyBase:FBgn0005563
 C:Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match 72.7%; Score 32; DB 2; Length 1851;
 Best Local Similarity 55.6%; Pred. No. 1.9e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
 Db 485 MSEMFLKIF 493

RESULT 9

A41098
 calcium channel protein alpha-1 chain isoform A - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 24-Sep-1999
 C:Accession: A41098; A35901
 R:Starr, T.V.B.; Pryce, W.; Snutch, T.P.
 Proc. Natl. Acad. Sci. U.S.A. 88, 5621-5625, 1991
 A>Title: Primary structure of a calcium channel that is highly expressed in the rat cere
 A:Reference number: A41098; MUID:91288516; PMID:164826
 A:Accession: A41098
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2212 <GBA>
 A:Cross-references: GB:M64373; NID:G203110; PIDN:AAA40806.1; PID:G203111
 R:Snutch, T.P.; Leonard, J.P.; Gilbert, M.M.; Lester, H.A.; Davidson, N.

Proc. Natl. Acad. Sci. U.S.A. 87, 3391-3395, 1990
A>Title: Rat brain expresses a heterogeneous family of calcium channels.
A/Reference number: A35901; PMID:90239020; PMID:1692134
A/Accession: A35901
A/Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A/Molecule type: mRNA
A/Residues: 1435-1667 <SNV>
C/Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match 72.7%; Score 32; DB 2; Length 2212;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
DB 533 MSBMTKMY 541

RESULT 10
146477
Calcium channel BI-1 - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 17-Nov-2000
C/Accession: 146477; 146478
R/Mol. Y.; Friedrich, T.; Kim, M.S.; Mikami, A.; Nakai, J.; Ruth, P.; Bosse, E.; Hofman
Nature 350, 398-402, 1991
A>Title: Primary structure and functional expression from complementary DNA of a brain c
A/Reference number: 146477; PMID:91187110; PMID:1849233
A/Accession: 146477
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-2273 <MOR>
A/Cross-references: EMBL:X57476; NID:g1522; PIDD:CAA40714.1; PID:g1523
A/Accession: 146478
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1856, 'R', 1860-1862, 'GV', 1865-1866, 'VIS', 1870-1876, 'K', 1878-1879, '
A/Cross-references: EMBL:X57688; NID:g1524; PIDD:CAA40871.1; PID:g1525
C/Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match 72.7%; Score 32; DB 2; Length 2273;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
DB 531 MSBMTKMY 539

RESULT 11
146480
Calcium channel BI-2 - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 17-Nov-2000
C/Accession: 146480; 146479
R/Mol. Y.; Friedrich, T.; Kim, M.S.; Mikami, A.; Nakai, J.; Ruth, P.; Bosse, E.; Hofman
Nature 350, 398-402, 1991
A>Title: Primary structure and functional expression from complementary DNA of a brain c
A/Reference number: 146477; PMID:91187110; PMID:1849233
A/Accession: 146480
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-2242 <MOR>
A/Cross-references: EMBL:X57689; NID:g1528; PIDD:CAA40872.1; PID:g1529
A/Accession: 146479
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1856, 'L', 1858, 'R', 1860-1862, 'MV', 1865-1866, 'HMP', 1870-1876, 'N', 1878-1879, '
A/Cross-references: EMBL:X57477; NID:g1526; PIDD:CAA40715.1; PID:g1527
C/Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match 72.7%; Score 32; DB 2; Length 2424;
Best Local Similarity 55.6%; Pred. No. 2.5e+02;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
DB 531 MSBMTKMY 539

RESULT 12
157007
Pregnancy-specific glycoprotein - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 02-Aug-1995 #sequence_revision 02-Aug-1996 #text_change 21-Jan-2000
C/Accession: 157007
R/Rudert, F.; Sanders, A.M.; Thompson, J.A.; Rebstock, S.; Zimmermann, W.A.
Mamm. Genome 3, 262-273, 1992
A>Title: Characterization of murine carcinoembryonic antigen gene family members.
A/Reference number: 157007; PMID:92345715; PMID:1638085
A/Accession: 157007
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-141 <RES>
A/Cross-references: GB:M83341; NID:g200309; PIDD:AAA39314.1; PID:g554253
C/Genetics:
A/Gene: CEM7
A/Insertion: 22/1
C/Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-term
C/Keywords: glycoprotein
F/137/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEA1>
F/48-96/Domain: immunoglobulin homology <IMM>

Query Match 70.5%; Score 31; DB 2; Length 141;
Best Local Similarity 62.5%; Pred. No. 20;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
DB 134 SEMFLQIY 141

RESULT 13
S23659
Superoxide dismutase (EC 1.15.1.1) (Mn) - European spiny lobster (fragment)
C/Species: Palinurus vulgaris (European spiny lobster)
C/Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C/Accession: S23659
R/Smith, M.W.; Doolittle, R.F.
J. Mol. Evol. 34, 175-184, 1992
A>Title: A comparison of evolutionary rates of the two major kinds of superoxide dismuta
A/Reference number: S23655; PMID:92211732; PMID:1556751
A/Accession: S23659
A/Molecule type: mRNA
A/Residues: 1-144 <SMI>
A/Cross-references: EMBL:X64063; NID:g10095; PIDD:CAA45419.1; PID:g1335662
C/Genetics:
A/Genome: nuclear
A/Function:
C/Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
C/Superfamily: superoxide dismutase (Mn)
C/Keywords: manganese; metalloprotein; mitochondrion; oxidoreductase

Query Match 70.5%; Score 31; DB 2; Length 144;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
DB 2 ISEMIMQIH 10

RESULT 14
Q08BNS
Integral membrane protein - saimiriine herpesvirus 1 (strain 11)
C/Species: saimiriine herpesvirus 1

Job time : 25 secs

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE, AND IDENTIFICATION BY MASS SPECTROMETRY.
RC STRAIN=CV, US-Hannamkigoenut;
RX MEDLINE=20212743; PubMed=10750705;
RA Hanai H., Matsuno T., Yamamoto M., Matsubayashi Y., Kobayashi T.,
RA Kamada H., Sakagami Y.;
RT "A secreted peptide growth factor, phytosulfokine, acting as a
stimulatory factor of carrot somatic embryo formation.";
PL Plant Cell Physiol. 41:27-32 (2000).
CC -1- FUNCTION: IN PRESENCE OF 2,4-D, STIMULATES PROLIFERATION OF THE
CELLS. BUT DOES NOT STIMULATE DIFFERENTIATION INTO THE SOMATIC
EMBRYOS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: SULFATION IS IMPORTANT FOR ACTIVITY AND FOR THE BINDING TO A
CC PUTATIVE MEMBRANE RECEPTOR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PHYTOSULFOKINE FAMILY.
KM Growth factor; Sulfation. PHYTOSULFOKINE-BETA.
FT PEPTIDE 1 4
FT MOD_RES 1 1 SULFATION.
FT MOD_RES 3 3 SULFATION.
SQ SEQUENCE 5 AA; 687 MW; 76C1BB504B300000 CRC64;

Query Match 25.0%; Score 11; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 1Y 9
DB 2 1Y 3

RESULT 15
RE21 LITRU STANDARD; PRT; 5 AA.
ID RE21 LITRU
AC P82071;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Rubellidin 2.1.
OS Litorea rubella (Desert tree frog).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Pelodyadinae; Litorea.
OX NCBI_TaxID=104895;
RN [3]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=Skin secretion;
RA Steinhorn S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT "The structure of new peptides from the Australian red tree frog
'Litorea rubella', the skin peptide profile as a probe for the study
of evolutionary trends of amphibians.";
PL Aust. J. Chem. 49:955-963 (1996).
CC -1- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTI-BIOTIC
ACTIVITY.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
CC -1- MASS SPECTROMETRY: MW=626; METHOD=PAB.
KW Amphibian skin.
SQ SEQUENCE 5 AA; 626 MW; 6DD9C9CB10300000 CRC64;

Query Match 25.0%; Score 11; DB 1; Length 5;
Best Local Similarity 66.7%; Pred. No. 1.1e+05;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 EMF 5
DB 2 ERF 4

Search completed: May 1, 2003, 23:24:03

A:Accession: A35175
 A:Molecule type: mRNA
 A:Residues: 1-952,1033-1344 <LIG1>
 A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA5804.1; PID:g182124; GB:
 A:Experimental source: splice form A
 A:Note: Genbank entries H0MEPIS1A and H0MEPIS1A2 present only the amino- and carboxyl-ter
 A:Accession: B35175
 A:Molecule type: mRNA
 A:Residues: 1-19,29-992,1033-1344 <LIG2>
 A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA5806.1; PID:g182129; GB:
 A:Experimental source: splice form B
 A:Note: Genbank entries H0MEPIS1A and H0MEPIS1A2 present only the amino- and carboxyl-ter
 J:Gendler, S.J.; Lamasera, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Pest, N.; Burchel
 J. Biol. Chem. 265, 15286-15293, 1990
 A:Title: Molecular cloning and expression of human tumor-associated polymorphic epithel
 A:Reference number: A35886; MUID:90368715; PMID:1697389
 A:Accession: A35886
 A:Molecule type: mRNA
 A:Status: not compared with conceptual translation
 A:Residues: 1-19,29-992,1033-1344 <GEN>
 A:Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870
 A:Note: Genbank entry HUMMUCAB includes one copy of the tandemly repeated sequence
 R:Lin, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
 J. Biol. Chem. 265, 15294-15299, 1990
 A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
 A:Reference number: A35887; MUID:90368716; PMID:2394722
 A:Accession: A35887
 A:Molecule type: mRNA
 A:Status: not compared with conceptual translation
 A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
 A:Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599
 A:Note: Genbank entry HUMPANMU contains four fewer copies of the tandemly repeated sequ
 R:Wreschner, D.H.; Harevenet, M.; Tatarfay, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
 Eur. J. Biochem. 189, 463-473, 1990
 A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
 A:Reference number: S10571; MUID:90276413; PMID:2351132
 A:Accession: S10572
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <MRB>
 A:Cross-references: EMBL:X52229; NID:g37053
 R:Wreschner, D.H.
 submitted to the EMBL Data Library, March 1990
 A:Reference number: S40293
 A:Accession: S40293
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <MRZ>
 A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA6478.1; PID:g37054
 R:Abu, M.; Siddiqui, U.; Kute, D.
 Biochem. Biophys. Res. Commun. 165, 644-649, 1989
 A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
 A:Reference number: A36735; MUID:90088473; PMID:2597151
 A:Accession: A36735
 A:Molecule type: mRNA
 A:Residues: 1-142,'O',144-162,'O',164-168 <ABE>
 A:Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA5757.1; PID:g181543
 R:Masuawa, Y.; Miyachi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
 J. Biochem. 112, 609-615, 1992
 A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglut
 A:Reference number: U02035; MUID:9313189; PMID:1478919
 A:Accession: U02035
 A:Molecule type: mRNA
 A:Residues: 998-1011,'E',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>
 A:Experimental source: gastric carcinoma cell
 R:Zirhan-Licht, S.; Barnuch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
 FEBS Lett. 356, 130-136, 1994
 A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytolokine
 A:Reference number: S51026; MUID:95080414; PMID:7989707
 A:Accession: S51026
 A:Molecule type: DNA
 A:Status: annotation
 A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region at
 C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
 partial repeats. The repeat shown is defined by small nucleic sites.
 C:Comment: Sdkline and theonine residues in the tandem repeat domain are extensively gly
 C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146

C:Genetics:
 A:Gene: GDB:MUC1; PUBM
 A:Cross-references: GDB:120705; OMIM:158340
 A:Map position: 1q23-1q23
 A:Intron: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
 C:Superfamily: polymorphic epithelial mucin
 C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphs
 F:1-62/Region: mucin 1 precursor, splice form A #status predicted <PRA>
 F:1-1344/Product: mucin 1 amino-terminal non-repetitive
 F:1-19,29-32/Domain: signal sequence #link PRA #status predicted <SIGA>
 F:1-19,29-1344/Product: signal sequence #link PRA #status predicted <SIGB>
 F:1-19,29-212,1033-1344/Product: mucin 1 precursor, splice form B #status predicted <PRB>
 F:138-1017/Region: 20-residue repeats (GSTAPPAGVTASAPDRPAP)
 F:143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
 F:1243-1272/Domain: transmembrane #status predicted <TM>
 F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted
 Query Match 100.0%; Score 44; DB 1; Length 1344;
 Best Local Similarity 100.0%; Pred. No. 0.29;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ISEMFLOLY 9
 Db 1162 ISEMFLOLY 1170
 RESULT 3
 S54293
 C:Regulator protein p122-RhoGAP - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 01-Sep-1995
 C:Accession: S54293
 R:Homma, Y.; Emori, Y.
 EMBO J. 14, 286-291, 1995
 A:Title: A dual functional signal mediator showing RhoGAP and phospholipase C-delta acti
 A:Reference number: S54293; MUID:95137008; PMID:783539
 A:Accession: S54293
 A:Molecule type: mRNA
 A:Status: preliminary; nucleic acid sequence not shown
 A:Residues: 1-1083 <ROM>
 A:Cross-references: EMBL:D31962
 Query Match 81.8%; Score 36; DB 2; Length 1083;
 Best Local Similarity 77.8%; Pred. No. 14;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 ISEMFLOLY 9
 Db 721 ISEMFLOLY 729
 RESULT 4
 G59435
 DUC-1 (deleted in liver cancer), p122 [imported] - human
 C:Species: Homo sapiens (man)
 C:Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 03-Jun-2002
 C:Accession: G59435
 R:Yuan, B.Z.; Miller, M.J.; Keck, C.L.; Zimonjic, D.B.; Thorgeirsson, S.S.; Popescu, N.
 Cancer Res. 58, 2196-2199, 1998
 A:Title: Cloning, characterization, and chromosomal localization of a gene frequently de
 A:Reference number: G59435
 A:Accession: G59435
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1091 <YUA>
 A:Cross-references: GB:NP_006085; PID:g5174427; PIDN:NP_006085.1
 Query Match 81.8%; Score 36; DB 2; Length 1091;
 Best Local Similarity 77.8%; Pred. No. 14;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:08:21 ; Search time 0.946429 Seconds
(without alignments)
914.184 Million cell updates/sec

Title: US-09-658-621B-26

Perfect score: 44

Sequence: 1 ISEMFLQIT 9

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 73: *
2: PIR1: *
3: PIR2: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	255	1	S48146
2	44	100.0	1344	1	A35175
3	36	81.8	1083	2	S54293
4	36	81.8	1091	2	G59435
5	33	75.0	333	2	D55283
6	33	75.0	381	2	T29826
7	33	75.0	2223	2	A47447
8	32	72.7	1851	2	T13980
9	32	72.7	2212	2	A41098
10	32	72.7	2273	2	T46477
11	32	72.7	2424	2	T46480
12	31	70.5	141	2	T57007
13	31	70.5	144	2	S23659
14	31	70.5	366	1	OQ8BNS
15	31	70.5	426	2	T41682
16	31	70.5	477	2	T04916
17	31	70.5	491	2	C70868
18	31	70.5	492	2	T23502
19	31	70.5	606	2	H64986
20	31	70.5	896	2	S76064
21	31	70.5	896	2	S59980
22	31	70.5	924	2	T04910
23	31	70.5	995	2	H59432
24	31	70.5	1071	2	S38164
25	31	70.5	1631	1	SAZQX1
26	31	70.5	1639	1	S05603
27	31	70.5	1640	2	A24594
28	31	70.5	2670	2	T37919
29	30	68.2	117	2	C71811

30	30	68.2	244	2	H69844	diadenosine tetrap
31	30	68.2	332	2	T41227	hypothetical prote
32	30	68.2	334	2	F69469	conserved hypothet
33	30	68.2	383	2	A13224	two component sens
34	30	68.2	385	2	T24328	hypothetical prote
35	30	68.2	390	2	D81289	probable aminotran
36	30	68.2	408	2	S76830	hypothetical prote
37	30	68.2	416	2	AF2070	hypothetical prote
38	30	68.2	579	2	AC2200	potassium-transport
39	30	68.2	630	2	A39344	tumor-associated m
40	30	68.2	631	2	T52257	epistatin - mouse
41	30	68.2	770	2	S37962	mitochondrial inte
42	30	68.2	815	2	S67675	probable membrane
43	30	68.2	946	2	A71843	D-lactate dehydrog
44	30	68.2	948	2	F64672	D-lactate dehydrog
45	30	68.2	981	2	S55132	hypothetical prote

ALIGNMENTS

RESULT 1

S48146
mucin 1 precursor, non-repetitive splice form Y [validated] - human
N/Alternate names: breast carcinoma-associated DF3 antigen; epistatin; MUC1/Y protein.
C/Species: Homo sapiens (man)
C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C/Accession: S48146
R/Zrhan-Licht, S.; Vos, H.L.; Baruch, A.; Elroy-Stein, O.; Sagiv, D.; Kaydar, I.; Hilk
Bur, J. Biochem. 224, 787-795, 1994
A/Title: Characterization and molecular cloning of a novel MUC1 protein, devoid of land
A/Reference number: S48146; PMID:95010060; PMID:7925397
A/Accession: S48146
A/Molecule type: mRNA
A/Residues: 1-255 <Zrhan>
A/Cross-reference: EMBL:X80761; NID:9541679; PIDN:CA56734.1; PID:9541680
A/Comment: For alternative splice forms with a tandem repeat domain, see PIR:A35175.

A/Genetic: GDB:MUCL; PUM
A/Cross-reference: GDB:120705; OMIM:158340
A/Map position: 1q21-1q23
A/Intons: 20/1; 53/3; 73/3; 95/2; 141/1; 181/3; 231/3
C/Suprafamily: polymorphic epithelial mucin
C/Keywords: glycoprotein; phosphoprotein; polymorphism; transmembrane protein
F/1-53/Region: mucin 1 amino-terminal non-repetitive
F/1-23/Domain: signal sequence #status predicted <Sig>
F/24-255/Product: mucin 1, non-repetitive splice form Y #status predicted <MAT>
F/154-255/Region: mucin 1 carboxyl-terminal non-repetitive
F/156-183/Domain: transmembrane #status predicted <TRM>
F/155,133/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 44; DB 1; Length 255;
Best local similarity 100.0%; Pred. No. 0.05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMFLQIT 9
DB 73 ISEMFLQIT 81

RESULT 2

A35175
mucin 1 precursor, repetitive splice form A [validated] - human
N/Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; epistatin
necrotic mucin; polymorphic epithelial mucin (PEM)
N/Content: mucin 1 precursor; epithelial tumor antigen splice form; mucin 1 precursor,
C/Species: Homo sapiens (man)
C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Jun-2000
C/Accession: A35175; B35175; A35867; A35867; S10572; S40293; A36735; PX0066; S10218; S51
R/Ligtenberg, M.J.L.; Vos, H.L.; Gemmessen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A/Title: Epistatin, a carcinoma-associated mucin, is generated by a polymorphic gene enc
A/Reference number: A35175; PMID:90202794; PMID:2318825

A>Note: host Saimiri sciureus (common squirrel monkey)
 C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
 C:Accession: A36810
 R:Albrecht, J.
 submitted to the EMBL Data Library, January 1992
 A:Description: Primary structure of the herpesvirus saimiri genome.
 A:Reference number: A36806
 A:Accession: A36810
 A:Molecule type: DNA
 A:Residues: 1-356 <ALB>
 A:Cross-references: GB:X64346; NID:G60320; PIDN:CA45662.1; PID:G60360
 R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.; W
 J:Virol. 66, 5047-5058, 1992
 A:Title: Primary structure of the herpesvirus saimiri genome.
 A:Reference number: A37309; WUID:92333688; PMID:1321287
 A:Contents: annotation; possible protein-coding frames
 A>Note: neither amino acid nor nucleotide sequence is given
 C:Genetics:
 A:Gene: 39
 C:Superfamily: cytomegalovirus UL100 protein
 C:Keywords: transmembrane protein
 F:18-34/Domain: transmembrane #status predicted <TM1>
 F:87-103/Domain: transmembrane #status predicted <TM2>
 F:152-168/Domain: transmembrane #status predicted <TM3>
 F:214-230/Domain: transmembrane #status predicted <TM4>
 F:236-252/Domain: transmembrane #status predicted <TM5>
 F:274-290/Domain: transmembrane #status predicted <TM6>
 F:305-321/Domain: transmembrane #status predicted <TM7>

Query Match 70.5%; Score 31; DB 1; Length 366;
 Best Local Similarity 66.7%; Pred. No. 55;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMPLOIY 9
 ::|||
 Db 259 MTEMFLODY 267

RESULT 15
 T41682
 hypothetical protein SPCP1E11.03 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
 C:Accession: T41682
 R:Wedler, H.; Duesterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, September 1999
 A:Reference number: Z21742
 A:Accession: T41682
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-426 <MED>
 A:Cross-references: EMBL:AL117183; PIDN:CA854862.1; GSPDB:GNO0068; SPDB:SPCP1E11.03
 A:Experimental source: strain 972h-; clone p1 p1E11
 C:Genetics:
 A:Gene: SPDB:SPCP1E11.03
 A:Map position: 3
 A:Introns: 217/2; 300/3; 411/2
 C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPCP1E11.03

Query Match 70.5%; Score 31; DB 2; Length 426;
 Best Local Similarity 66.7%; Pred. No. 65;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ISEMPLOIY 9
 ::|||
 Db 127 IKEMFLDFY 135

Search completed: May 1, 2003, 23:13:53
 Job time : 2.94643 secs

Thu May 8 16:14:19 2003

us-09-658-621b-26.rspt

Page 6

Db :||||:|
533 MSEMFKMY 541

Search completed: May 1, 2003, 23:12:50
Job time : 4.91071 secs

Query Match 75.0%; Score 33; DB 10; Length 955;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
Db 920 SQUFLQVY 927

RESULT 12

Q9EN19 PRELIMINARY; PRT; 290 AA.
AC Q9EN19;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE AMV029.
GN Amasacta moorei entomopoxvirus (AmEPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=28321;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20396580; PubMed=10936094;
RA Bawden A.L., Glasberg K.J., Diggins J., Shaw R., Farmerie W.,
Moyer R.W.;
RT "Complete Genomic Sequence of the Amasacta moorei Entomopoxvirus:
RT Analysis and Comparison with Other Poxviruses.";
RT Virology 274:120-139 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Bawden A.L., Glasberg K.J., Diggins J., Shaw R., Farmerie W.,
Moyer R.W.;
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF250284; AAG02735.1; -;
DR InterPro; IPR001810; P-box.
DR PROSITE; PS50181; FBOX; 1.
SQ SEQUENCE 290 AA; 35839 MW; 709D59A7F0DB8C79 CRC64;

Query Match 72.7%; Score 32; DB 12; Length 290;
Best Local Similarity 55.6%; Pred. No. 64;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMPLQIY 9
Db 207 LSQFLQIY 215

RESULT 13

Q9JG85 PRELIMINARY; PRT; 765 AA.
AC Q9JG85;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE PORF1.
GN TT virus.
OC Viruses; ssDNA viruses, unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20417334; PubMed=10963344;
RA Tanaka Y., Orito E., Onno T., Nakano T., Hayashi K., Kato T.,
RA Mukaide M., Iida S., Mizokami M.;
RT "Identification of a 23kDa protein encoded by putative open reading
RT frame 2 of TT virus (TTV) genotype 1 different from the other
RT genotypes.";
RL Arch. Virol. 145:1385-1398 (2000).
DR EMBL; AB030488; BAA90409.1; -;
DR InterPro; IPR004219; TTVirus_Unk.
DR Pfam; PF02956; TT_ORF1; 1.
SQ SEQUENCE 765 AA; 90275 MW; 707029EB45829B5F CRC64;

Query Match 72.7%; Score 32; DB 12; Length 765;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMPLQIY 9
Db 731 TREMFLOTH 739

RESULT 14

Q9USY1 PRELIMINARY; PRT; 2015 AA.
AC Q9USY1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Microtubule-associated protein CP224.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RX MEDLINE=20233854; PubMed=10769206;
RA Graef R., Daubert C., Schliwa M.;
RT "Dictyostelium DdCP224 is a microtubule-associated protein and a
RT permanent centrosomal resident involved in centrosome duplication.";
RT J. Cell Sci. 113:1747-1758 (2000).
DR EMBL; AJ012088; CAB56504.1; -;
DR InterPro; IPR000357; HEAT repeat.
DR PROSITE; PS50077; HEAT_REPEAT; 1.
SQ SEQUENCE 2015 AA; 224041 MW; BA64B982MDC922EE CRC64;

Query Match 72.7%; Score 32; DB 5; Length 2015;
Best Local Similarity 62.5%; Pred. No. 3.8e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
Db 184 SELFLEIY 191

RESULT 15

Q8RSM7 PRELIMINARY; PRT; 2327 AA.
AC Q8RSM7;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CAV2.1.
GN CACNALA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6N;
RX MEDLINE=21850703; PubMed=11756409;
RA Tanemura T., Saegusa H., Ishikawa K., Nagayama S., Murakoshi T.,
RA Mizusawa H., Tanabe T.;
RT "Novel Cav2.1 Splice Variants Isolated from Purkinje Cells Do Not
RT Generate P-type Ca²⁺ Current.";
RL U. Biol. Chem. 277:7214-7221 (2002).
DR EMBL; AB066608; BAB85611.1; -;
SQ SEQUENCE 2327 AA; 263346 MW; 21FMA3297E7893C2 CRC64;

Query Match 72.7%; Score 32; DB 11; Length 2327;
Best Local Similarity 55.6%; Pred. No. 4.4e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMPLQIY 9

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:07:31 ; Search time 1.91071 Seconds
(without alignments)
970.540 Million cell updates/sec

Title: US-09-658-621b-26
Perfect score: 44
Sequence: 1 ISEMFLQIY 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phase:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriaph:*
17: SP archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	93.2	553	6 Q9MZL1	Q9MZL1 macaca mla
2	36	81.8	1083	11 Q63744	Q63744 rattus ratc
3	36	81.8	1084	11 Q8R541	Q8R541 mus musculu
4	36	81.8	1091	4 Q43199	Q43199 homo sapien
5	36	81.8	1091	4 Q96GB1	Q96GB1 homo sapien
6	36	81.8	1554	4 Q9CGB0	Q9CGB0 homo sapien
7	33	75.0	182	10 Q8RX11	Q8RX11 arabidopsis
8	33	75.0	193	11 Q6Q408	Q6Q408 cavia cutie
9	33	75.0	330	5 Q9UJG3	Q9UJG3 caenorhabd
10	33	75.0	333	16 Q930M3	Q930M3 rhizobium m
11	33	75.0	955	10 Q458S5	Q458S5 arabidopsis
12	32	72.7	290	12 Q9EN19	Q9EN19 amesacta moo
13	32	72.7	765	12 Q9JGS5	Q9JGS5 et virus. p
14	32	72.7	2015	5 Q9U5Y1	Q9U5Y1 dictyostel
15	32	72.7	2327	11 Q8R5M7	Q8R5M7 mus musculu
16	32	72.7	2365	11 Q8R5M6	Q8R5M6 mus musculu

17	32	72.7	2472	4 Q9NS89	Q9NS89 homo sapien
18	32	72.7	2506	4 Q9NS88	Q9NS88 homo sapien
19	31	70.5	119	11 Q08928	Q08928 mus musculu
20	31	70.5	277	16 Q8RA94	Q8RA94 thermoaer
21	31	70.5	426	3 Q9U84	Q9U84 schizosacch
22	31	70.5	437	5 Q8PWP6	Q8PWP6 paramet
23	31	70.5	450	17 Q8TME0	Q8TME0 mechanosar
24	31	70.5	455	11 P70161	P70161 mus musculu
25	31	70.5	477	10 Q49640	Q49640 arabidopsis
26	31	70.5	492	5 Q93867	Q93867 caenorhabd
27	31	70.5	510	10 Q8W024	Q8W024 arabidopsis
28	31	70.5	599	10 Q8W4D7	Q8W4D7 arabidopsis
29	31	70.5	605	17 Q8SFU8	Q8SFU8 arabidopsis
30	31	70.5	612	17 Q8TND2	Q8TND2 mechanosar
31	31	70.5	697	5 Q9N632	Q9N632 leishmania
32	31	70.5	896	16 Q55544	Q55544 synchocyst
33	31	70.5	974	10 Q49634	Q49634 arabidopsis
34	31	70.5	995	4 Q9Y3M8	Q9Y3M8 homo sapien
35	31	70.5	1056	11 Q923Q2	Q923Q2 mus musculu
36	31	70.5	1326	5 Q9W1Q5	Q9W1Q5 drosophila
37	30	68.2	117	16 Q92JB0	Q92JB0 helicobacte
38	30	68.2	141	11 Q9D740	Q9D740 mus musculu
39	30	68.2	203	5 Q76189	Q76189 trypanosoma
40	30	68.2	235	12 Q82936	Q82936 jerry sloug
41	30	68.2	235	12 Q98669	Q98669 san angelo
42	30	68.2	235	12 Q89796	Q89796 jamescown c
43	30	68.2	244	16 Q31614	Q31614 bacillus su
44	30	68.2	277	2 Q55036	Q55036 synchocyst
45	30	68.2	332	3 Q94472	Q94472 schizosacch

ALIGNMENTS

RESULT 1	Q9MZL1	PRELIMINARY;	PRT;	553 AA.
ID	Q9MZL1			
AC	Q9MZL1			
DT	01-OCT-2000 (TREMBL)	15, Created		
DT	01-OCT-2000 (TREMREL)	15, Last sequence update		
DT	01-MAR-2002 (TREMREL)	20, Last annotation update		
DE	Mucin 1 (Fragment).			
OS	Macaca mulatta (Rhesus macaque).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;			
OC	Cercopithecoidea; Macaca.			
OX	NCBI_Taxid=9544;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20330533; PubMed=1086975;			
RA	Vaughan H.A., Ho D.W.M., Karanikas V., Sandrin M.S., McKenzie I.F.C.,			
RT	Pietersz G.A.,			
RT	"The Immune Response of Mice and Cynomolgus Monkeys to Macaque Mucin1-			
RT	Mannan."/			
RL	Vaccine 18:3297-3309(2000).			
DR	EMBL; AF16947; AAF82403.1; "			
DR	InterPro; IPR001064; Crystalin.			
DR	InterPro; IPR000082; SEA_domain.			
DR	PFam; PF01390; SEA; 1.			
DR	SMART; SM00200; SEA; 1.			
DR	PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.			
DR	PROSITE; PS50024; SEA; 1.			
FT	NON_TER	1		
FT	NON_TER	1		
FT	NON_TER	1		
SQ	SEQUENCE	553 AA;	55778 MW;	60786DD2E829318 CRC64;
Query Match		93.2%;	Score 41;	DB 6;
Best Local Similarity		88.9%;	Pred. No. 1.3;	
Matches	8;	Conservative	1;	Mismatches
			0;	Indels
			0;	Gaps
Qy	1 ISEMFLQIY 9			
Db	396 ISEFLQIY 404			

RESULT 2

Q63744

ID 063744 PRELIMINARY; PRT; 1083 AA.

AC 063744;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE RhogAP.

OS Rattus rattus (Black rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10117;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RX MEDLINE=95137008; PubMed=7835339;

RT "A dual functional signal mediator showing RhogAP and phospholipase C-

delta stimulating activities";

RL EMBL J.14:286-291(1995).

DR EMBL; D31962; BAA21675.1; -.

DR HSSP; Q07960; IRGP.

DR InterPro; IPR001005; Myb_DNA_binding.

DR InterPro; IPR00198; RhGAP.

DR InterPro; IPR001660; SAM.

DR InterPro; IPR002913; START.

DR Pfam; PF00620; RhogAP.1.

DR Pfam; PF01852; START.1.

DR SMART; SM00324; RhogAP.1.

DR SMART; SM00454; SAM.1.

DR SMART; SM00234; START.1.

DR PROSITE; PS00037; MYB.1; UNKNOWN.1.

SQ SEQUENCE 1083 AA; 122465 MW; 076DDE77ACD9D012 CRC64;

Query Match

Best Local Similarity 81.8%; Score 36; DB 11; Length 1083;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9

Db 721 ISETFLQIY 729

RESULT 3

Q6R541

ID 06R541 PRELIMINARY; PRT; 1084 AA.

AC 06R541;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE DIC-1.

GN ARHGAP7.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SV;

RX MEDLINE=20169180; PubMed=10702663;

RA Yuan B.Z., Yang Y., Keck-Waggoner C.L., Zimonjic D.B.,

RA Thorngelirsson S.S., Popescu N.C.;

RA "Assignment and cloning of mouse Arhgap7 to chromosome 8A4-B2, a

conserved syntenic region of human chromosome 9p22--p21.";

RT Cytogenet. Cell Genet. 87:189-190(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SV;

RA Durlin M.E., Yuan B.Z., Thorngelirsson S.S., Popescu N.C.;

RA "Gene structure, tissue expression and linkage mapping of the mouse

DIC-1 gene (Arhgap7).";

RL Gene 0:0-0(2002).

DR EMBL; AF411442; AAL87620.1; -.

DR EMBL; AF411435; AAL87620.1; JOINED.

DR EMBL; AF411436; AAL87620.1; JOINED.

DR EMBL; AF411437; AAL87620.1; JOINED.

DR EMBL; AF411438; AAL87620.1; JOINED.

DR EMBL; AF411439; AAL87620.1; JOINED.

DR EMBL; AF411440; AAL87620.1; JOINED.

DR EMBL; AF411441; AAL87620.1; JOINED.

SQ SEQUENCE 1084 AA; 122418 MW; 9FB1B4176ACABDEB CRC64;

Query Match

Best Local Similarity 81.8%; Score 36; DB 11; Length 1084;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9

Db 722 ISETFLQIY 730

RESULT 4

Q43199

ID 043199 PRELIMINARY; PRT; 1091 AA.

AC 043199;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Deleted in liver cancer-1.

GN DIC-1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Yuan B.Z., Miller M.J., Keck C.L., Zimonjic D.B., Thorngelirsson S.S.,

RA Popescu N.C.;

RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF035119; AAB87700.1; -.

DR HSSP; Q07960; IRGP.

DR InterPro; IPR001005; Myb_DNA_binding.

DR InterPro; IPR00198; RhogAP.

DR InterPro; IPR001660; SAM.

DR InterPro; IPR002913; START.

DR Pfam; PF00620; RhogAP.1.

DR Pfam; PF01852; START.1.

DR SMART; SM00324; RhogAP.1.

DR SMART; SM00454; SAM.1.

DR SMART; SM00234; START.1.

DR PROSITE; PS00037; MYB.1; UNKNOWN.1.

SQ SEQUENCE 1091 AA; 122816 MW; 51712DE7ECDD0F52A CRC64;

Query Match

Best Local Similarity 81.8%; Score 36; DB 4; Length 1091;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9

Db 729 ISETFLQIY 737

RESULT 5

Q960B1

ID 0960B1 PRELIMINARY; PRT; 1091 AA.

AC 0960B1; 014868; -.

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Deleted in liver cancer 1 (HP protein).

GN DIC1 OR HP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Jeong S.-J., Ditchchev A., Lerman M., Ditchchilo A., Jung M.;
RT "Identification of HP/DLCI exon and introns.";
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 9-1091 FROM N.A.
RC TISSUE=LUNG;
RA Wei M.-H., Pack S., Ivanov S., Lerman M.I.;
RT "Cloning and Molecular Characterization of the Human Ortholog of the
RT Rat Dual Regulator p12RhoGAP.";
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF408781; AAK97501.1; -.
DR EMBL; AF408768; AAK97501.1; JOINED.
DR EMBL; AF408770; AAK97501.1; JOINED.
DR EMBL; AF408772; AAK97501.1; JOINED.
DR EMBL; AF408774; AAK97501.1; JOINED.
DR EMBL; AF408775; AAK97501.1; JOINED.
DR EMBL; AF408776; AAK97501.1; JOINED.
DR EMBL; AF408777; AAK97501.1; JOINED.
DR EMBL; AF408778; AAK97501.1; JOINED.
DR EMBL; AF408779; AAK97501.1; JOINED.
DR EMBL; AF408780; AAK97501.1; JOINED.
DR EMBL; AF026219; AAB81637.1; -.
DR HSSP; Q07960; IRGP.
DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR000198; RhoGAP.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR002913; START.
DR Pfam; PF00620; RhoGAP; 1.
DR Pfam; PF01852; START; 1.
DR SMART; SM00324; RhoGAP; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00234; START; 1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
SQ SEQUENCE 1091 AA; 122858 MW; BF9CD0B20A80AE83 CRC64;

Query Match
Best Local Similarity 81.8%; Score 36; DB 4; Length 1091;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISEMPLOY 9
Db 729 LSEFLOY 737

RESULT 6
O9C0E0 PRELIMINARY; PRT; 1554 AA.
AC O9C0E0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA1723 protein (Fragment).
GN KIAA1723.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=1082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
DR EMBL; AB051510; BAB21814.1; -.
DR HSSP; Q07960; IRGP.

```

```

DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR000198; RhoGAP.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR002913; START.
DR Pfam; PF00620; RhoGAP; 1.
DR Pfam; PF01852; START; 1.
DR SMART; SM00324; RhoGAP; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00234; START; 1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
FT NON TER 1
SQ SEQUENCE 1554 AA; 173549 MW; 76FD31F139F2E12 CRC64;

Query Match
Best Local Similarity 81.8%; Score 36; DB 4; Length 1554;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISEMPLOY 9
Db 1192 LSEFLOY 1200

RESULT 7
O8RXL1 PRELIMINARY; PRT; 182 AA.
AC O8RXL1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 20.4 kDa protein (Fragment).
GN ATG44750.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.W., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Cariminci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.-J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.; Full length cDNA Clones.
RT "Arabidopsis Full length cDNA Clones."
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY080825; AAL87302.1; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 182 AA; 20423 MW; 05160C0CFC3B562 CRC64;

Query Match
Best Local Similarity 75.0%; Score 33; DB 10; Length 182;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 SEMFLOY 9
Db 153 SOLFLOY 160

RESULT 8
O60408 PRELIMINARY; PRT; 193 AA.
AC O60408;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Epithelial mucin (Fragment).
GN MUC1.
OS Cavia culteri (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.

```

```

OX NCBI_TaxID=10144;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ACTATING MAMMARY GLAND;
RX MEDLINE=96351712; PubMed=8747930;
RT Spicer A.P., Dunig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains.";
RL Mamm. Genome 6:885-888(1995).
DR EMBL; L41546; AAB48542.1; -.
DR InterPro: IPR000082; SEA_domain.
DR Pfam; PF03390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON TER
SQ SEQUENCE 193 AA; 21661 MW; D2593E4B9FEC12F CRC64;

Query Match 75.0%; Score 33; DB 11; Length 193;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
   :|||
   :|||
Db 14 ITRLFLQIY 22

RESULT 9
QY03G3 PRELIMINARY; PRT; 330 AA.
AC QY03G3.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE F40D4.2 protein.
F40D4.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhaditida; Rhaditoidae;
OC Rhaditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RL Matthews L.;
RN Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81536; CAB63199.1; -.
DR InterPro; IPR003003; 7TM_chemo2.
DR InterPro; IPR000168; 7TM_nematode.
DR Pfam; PF01604; 7tm 5; 1.
SQ SEQUENCE 330 AA; 37476 MW; C9B131FF6887A7AF CRC64;

Query Match 75.0%; Score 33; DB 5; Length 330;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQI 8
   :|||
   :|||
Db 208 VSGMFLQI 215

RESULT 10
QY03OM3 PRELIMINARY; PRT; 333 AA.
AC QY03OM3.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative Arac-type regulator.

```

```

GN RA0172 OR SMA0319.
OS Rhizobium meliloti (Sinorhizobium meliloti).
CG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fleher R.F., Jones T., Komp C., Abola A.P.,
RA Barlet-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL; AE007210; AAK64830.1; -.
DR InterPro; IPR000005; HTHArac.
DR Pfam; PF00165; HTH_Arac; 2.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; UNKNOWN_2.
DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
KM Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 333 AA; 36935 MW; 8BB2D702BD58A500 CRC64;

Query Match 75.0%; Score 33; DB 16; Length 333;
Best Local Similarity 62.5%; Pred. No. 44;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
   :|||
   :|||
Db 65 SDFVQVY 72

RESULT 11
QY0585 PRELIMINARY; PRT; 955 AA.
AC QY0585.
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein (SIMILARITY to DNA-damage-inducible protein
DE P).
GN T19K24.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA Sykes S.M., Kaul S., Mason T.M., Kervlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=99156233; PubMed=10048486;
RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
RT Sequence features of the regions of 1,081,958 bp covered by seventeen
RT physically assigned P1 and TAC clones.";
RL DNA Res 5:379-391(1998).
DR EMBL; AC002342; AAC79145.1; -.
DR EMBL; AB016874; BAB08828.1; -.
DR InterPro; IPR001126; UMCUC_1like.
DR Pfam; PF00817; IMS; 1.
SQ SEQUENCE 955 AA; 104984 MW; BF01A0B417AAA355 CRC64;

```

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:19:16 ; Search time 24 Seconds
(without alignments)
15.554 Million cell updates/sec

Title: US-09-658-621b-26
Perfect score: 44
Sequence: 1 ISEMFLQIV 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: * 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 231

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	34.1	6	1	CPD1_MYTED
2	15	34.1	8	1	CPD1_ENTPA
3	14	31.8	8	1	AL18_CARMA
4	13	29.5	9	1	ULAE_HUMAN
5	12	27.3	5	1	E104_LITRU
6	12	27.3	5	1	UC22_MALZE
7	12	27.3	6	1	CIP2_MYTED
8	12	27.3	6	1	E101_LITRU
9	12	27.3	8	1	AL16_CYPPO
10	12	27.3	9	1	PAR3_CALVO
11	12	27.3	9	1	FLNF_HIRME
12	11	25.0	4	1	FLRN_AMEL
13	11	25.0	5	1	PSK_DAVCA
14	11	25.0	5	1	RE21_LITRU
15	11	25.0	5	1	RE31_LITRU
16	11	25.0	7	1	E105_LITRU
17	11	25.0	7	1	PAR1_HELTI
18	11	25.0	7	1	PAR1_PROCL
19	11	25.0	7	1	PAR2_PROCL
20	11	25.0	7	1	HY_PIG
21	11	25.0	8	1	ANG2_BOTJA
22	11	25.0	8	1	PAR1_PANRE
23	11	25.0	8	1	PAR1_PANRE
24	11	25.0	8	1	PAR4_HOMAM
25	11	25.0	8	1	UPAA_HOMAM
26	11	25.0	8	1	PAR2_PANRE
27	11	25.0	9	1	PAR2_PANRE
28	11	25.0	9	1	FLN2_TREH
29	11	25.0	9	1	FLN2_TREH
30	11	25.0	9	1	UGAR_DIAAB
31	11	25.0	9	1	UGAR_DIAAB
32	10	22.7	4	1	FYRI_AMEL
33	10	22.7	7	1	LANC_CARUL

34	10	22.7	7	1	MMAL_ACHFU	P35919	achalina fu
35	10	22.7	7	1	MMAL_ACHFU	P35921	achalina fu
36	10	22.7	8	1	AL12_CARMA	P81815	carcinus ma
37	10	22.7	8	1	NPB_BOVIN	P15507	bos taurus
38	10	22.7	9	1	KOSF_GLYDA	P19853	clypeaster
39	10	22.7	9	1	NEUD_CAVPO	P34966	cavia porce
40	10	22.7	9	1	NEUD_CAVPO	P04277	homo sapien
41	10	22.7	9	1	OXYT_BUPRE	P42995	bufo. reguila
42	10	22.7	9	1	OXYT_CYPECA	P23879	cyprius ca
43	10	22.7	9	1	OXYT_RABIT	P32878	oryctolagus
44	10	22.7	9	1	OXYV_SQUAC	P43000	squalus aca
45	10	22.7	9	1	UN19_CLOPA	P81355	clostridium

ALIGNMENTS

RESULT 1
ID CPD1_MYTED STANDARD; PRT; 6 AA.
AC P13736;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Contraction-inhibiting peptide I (MIP I).
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxId=6550;
RN (1)
RP SEQUENCE.
RC TISSUE=pedal ganglion;
RX MEDLINE=88240357; PubMed=3377776;
RA Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.,
RT "Structures and actions of Mytilus inhibitory peptides.";
RL Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC MUSCLES.
CC -!- SIMILARITY: TO MIP II.
CC PIR; A27696; A27696.
DR Hormone; Amidation.
KW MOD RES
SQ SEQUENCE 6 AA; 637 MW; 72C9C6875B81000 CRC64;
Query Match 34.1%; Score 15; DB 1; Length 6;
Best Local Similarity 60.0%; Pred. No. 1;le+05;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 SEMFL 6
Db 2 SPMFV 6
RESULT 2
ID CPD1_ENTPA STANDARD; PRT; 8 AA.
AC P13269;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Sex pheromone CPD1.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxId=1351;
RN (1)
RP SEQUENCE.
RX MEDLINE=85040388; PubMed=6436978;
RA Suzuki A., Mori M., Sakagami Y., Isogai A., Fujino M., Kitada C.,
RT Craig R.A., Clewell D.B.;
RL "Isolation and structure of bacterial sex pheromone, cpd1.";
RL Science 226:849-850(1984).
CC -!- FUNCTION: CPD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
BACTERIOPHAGE PLASMID PPD1.

KM Pheromone. 8 AA; 913 MW; 8665B729C662C729 CRC64;
SQ SEQUENCE

Query Match 34.1%; Score 15; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 MFL 6
|||
DB 4 MFL 6

RESULT 3

AL18_CARMA STANDARD; PRT; 8 AA.
ID AL18_CARMA
AC P81821;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Carcinus maenas (Common shore crab) (Green crab).
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
OC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata;
OC Brachyura; Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RA MEDLINE=9812193; PubMed=9461295;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
"Isolation and identification of multiple neuropeptides of the
allatostatin superfamily in the shore crab Carcinus maenas";
RL Ent. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Amidation; Multigene family.
FT MOD RES 8 8
SQ SEQUENCE 8 AA; 919 MW; C82879D5B569AB5 CRC64;

Query Match 31.8%; Score 14; DB 1; Length 8;
Best Local Similarity 50.0%; Pred. No. 1.1e+05;
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 SEMF 5
|||
DB 1 SDMY 4

RESULT 4

U1AE_HUMAN STANDARD; PRT; 9 AA.
ID U1AE_HUMAN
AC P31931;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Unknown protein from 2D-page of liver tissue (Spot 115) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RC TISSUE=liver;
RA MEDLINE=9417969; PubMed=8313870;
RA Hughes G.J., Frutiger S., Paquet N., Pasquali C., Sanchez J.-C.,
Tissot J.-D., Balroch A., Appel R.D., Hochstrasser D.F.,
"Human liver protein map: update 1993";
RL Electrophoresis 14:1216-1222(1993).
CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 5.5, ITS MW IS: 11 KDa.
DR SWISS-2DPAGE; P31931; HUMAN.
FT NON_TER 9 9

SQ SEQUENCE 9 AA; 1096 MW; 477B3B173AE729C7 CRC64;

Query Match 29.5%; Score 13; DB 1; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.1e+05;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 4 MFL 8
|||
DB 2 LFLXL 6

RESULT 5

E104_LITRU STANDARD; PRT; 5 AA.
ID E104_LITRU
AC P82100;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Electrin 4.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Pelodidae; Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE.
RC TISSUE=Skin secretion;
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.,
"Peptides from the skin glands of the Australian buzzing tree frog
Litoria electrica. Comparison with the skin peptides from Litoria
rubella";
RL Aust. J. Chem. 52:639-645(1999).
CC -1- SUBCELLULAR LOCATION: SECRETED.
KW Amphibian skin; Amidation.
FT MOD RES 5 5
SQ SEQUENCE 5 AA; 616 MW; 61F2D1A059A00000 CRC64;

Query Match 27.3%; Score 12; DB 1; Length 5;
Best Local Similarity 20.0%; Pred. No. 1.1e+05;
Matches 1; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 5 FLOIY 9
|||
DB 1 FIVIV 5

RESULT 6

UC22_MAIZE STANDARD; PRT; 5 AA.
ID UC22_MAIZE
AC P80628;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Unknown protein from 2D-page of etiolated coleoptile (Spot 474) (Fragment).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE.
RC TISSUE=Coleoptile;
RA Touzet P., Riccardi F., Morin C., Damerval C., Huet J.-C.,
Pernollet J.-C., Zivy M., de Vienne D.,
"The maize two dimensional gel protein database: towards an integrated
genome analysis program";
RL Theor. Appl. Genet. 93:997-1005(1996).
CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 6.1, ITS MW IS: 30.4 KDa.
DR Maize-2DPAGE; P80628; COLEOPTILE.
FT NON_TER 1 1

ON	NCEI_TaxID=30608;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=JP285, JP288, JP289, JP292, JP308, AND JP313;
RX	MEDLINE=21184272; PubMed=11286490;
RA	Pastorini J., Martin R.D., Ehrmann P., Zimmermann E., Forstner M.R.;
RT	Molecular phylogeny of the Lemur family Cheirogaleidae (primates)
RL	based on mitochondrial DNA sequences."
RM	Mol. Phylogenet. Evol. 19:45-56(2001).
DR	EMBL; AF224624; AAK70547.1; -
DR	EMBL; AF224625; AAK70551.1; -
DR	EMBL; AF224626; AAK70555.1; -
DR	EMBL; AF224627; AAK70559.1; -
DR	EMBL; AF224628; AAK70563.1; -
DR	EMBL; AF224629; AAK70567.1; -
KM	Mitochondrion.
FT	NON_TER
SQ	SEQUENCE 9 AA; 1160 MW; DSC563636B5045A2 CRC64;
OY	5 FLOXY 9
Dy	:: ::
Dd	1 YVSIV 5
RESULT 7	
O94NB1	FRELIMITARY; PRT; 9 AA.
ID O94NB1	
AC O94NB1	
DT 01-DEC-2001	(TREMBlrel. 19, Created)
DT 01-DEC-2001	(TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001	(TREMBlrel. 19, Last annotation update)
DE	Cytochrome oxidase subunit III (Fragment).
GN	COIII
OS	Microcebus ravelobensis.
OC	Mitochondrion.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Strepsirrhini; Cheirogaleidae;
OC	Microcebus.
OX	NCEI_TaxID=122231;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=JP299, AND JP301;
RX	MEDLINE=21184272; PubMed=11286490;
RA	Pastorini J., Martin R.D., Ehrmann P., Zimmermann E., Forstner M.R.;
RT	"Molecular phylogeny of the Lemur family Cheirogaleidae (primates)
RL	based on mitochondrial DNA sequences."
RM	Mol. Phylogenet. Evol. 19:45-56(2001).
DR	EMBL; AF224630; AAK70571.1; -
DR	EMBL; AF224631; AAK70575.1; -
KM	Mitochondrion.
FT	NON_TER
SQ	SEQUENCE 9 AA; 1160 MW; DSC563636B5045A2 CRC64;
OY	5 FLOXY 9
Dy	:: ::
Dd	1 YVSIV 5
RESULT 8	
O94NB0	PRELIMINARY; PRT; 9 AA.
ID O94NB0	
AC O94NB0	
DT 01-DEC-2001	(TREMBlrel. 19, Created)
DT 01-DEC-2001	(TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001	(TREMBlrel. 19, Last annotation update)

```

DE Cytochrome oxidase subunit III (Fragment).
GN COIII.
OS Microcephus rufus (brown mouse lemur).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirhini; Cheirogaleidae;
OC Microcebus.
OX NCBI_TaxID=122232;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JP309, JP315, JP316, AND JP317;
RA MEDLINE=21184272; PubMed=11286490;
RA Pastorini J., Martin R.D., Ehresmann P., Zimmermann E., Forstner M.R.;
RT "Molecular phylogeny of the lemur family Cheirogaleidae (primates)
  based on mitochondrial DNA sequences."
RL Mol. Phylogenet. Evol. 19:45-56(2001).
RU Mol. Phylogenet. Evol. 19:45-56(2001).
DR EMBL; AF224636; AAK70595.1; -
DR EMBL; AF224637; AAK70599.1; -
DR EMBL; AF224638; AAK70603.1; -
DR EMBL; AF224639; AAK70607.1; -
KM Mitochondrion.
FT NON TER
SO SPOUDENCE 9 AA; 1160 MW; D5C563636S045A2 CRC64;

Query Match 34.1%; Score 15; DB 8; Length 9;
Best Local Similarity 40.0%; Pred. No. 6.7e+05;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLOQY 9
DB 1 YVSIV 5

RESULT 9
Q94NA9 PRELIMINARY; PRI; 9 AA.
ID Q94NA9; AC
Q94NA9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Cytochrome oxidase subunit III (Fragment).
GN COIII.
OS Daubentonia madagascariensis (Aye-aye).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirhini; Daubentonidae;
OC Daubentonia.
CX NCBI_TaxID=31869;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UP119, AND JP120;
RA MEDLINE=21184272; PubMed=11286490;
RA Pastorini J., Martin R.D., Ehresmann P., Zimmermann E., Forstner M.R.;
RT "Molecular phylogeny of the lemur family Cheirogaleidae (primates)
  based on mitochondrial DNA sequences."
RL Mol. Phylogenet. Evol. 19:45-56(2001).
RU Mol. Phylogenet. Evol. 19:45-56(2001).
DR EMBL; AF224641; AAK70615.1; -
DR EMBL; AF224642; AAK70619.1; -
KM Mitochondrion.
FT NON TER
SO SPOUDENCE 9 AA; 1160 MW; D5C563636S045A2 CRC64;

Query Match 34.1%; Score 15; DB 8; Length 9;
Best Local Similarity 40.0%; Pred. No. 6.7e+05;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLOQY 9
DB 1 YVSIV 5

RESULT 10
5050556

```

ID 05056 PRELIMINARY; PRT; 7 AA.
 AC 05056;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE GlyA (Fragment).
 GN GLYA.
 OS Actinobacillus actinomycetemcomitans (Haemophilus
 actinomycetemcomitans)
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Actinobacillus.
 OK NCBI_TaxID=714;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33384;
 RX MEDLINE=96355846; PubMed=8751884;
 RA Kojdubetz D., Siltzangel J. Jr., Wang B., Phillips L.H., Jacobs C.,
 Kraig E.;
 RT "cis Elements and trans factors are both important in strain-specific
 regulation of the leukotoxin gene in Actinobacillus
 actinomycetemcomitans."
 RL Intec. Immun. 64:3451-3460(1996).
 DR EMBL, U51862; AAB8721.1; -.
 FT NON TER 1
 SQ SEQUENCE 7 AA; 832 MW; 6DCB42D767340420 CRC64;

Query Match 29.5%; Score 13; DB 2; Length 7;
 Best Local Similarity 50.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 6 LQTY 9
 DB 3 LPVY 6

RESULT 11
 ID 068485 PRELIMINARY; PRT; 8 AA.
 AC 068485;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Aminoglycoside 3'-O-nucleotidyltransferase (Fragment).
 GN AANA1.
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Klebsiella.
 OK NCBI_TaxID=573;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1.
 RX MEDLINE=96287600; PubMed=9624504;
 RT "Characterization of the 6'-N-aminoglycoside acetyltransferase gene
 aac(6')-Iq from the integron of a natural multiresistance plasmid".
 RL Antimicrob. Agents Chemother. 42:1506-1508(1998).
 DR EMBL, AF047556; AAC25501.1; -.
 KW Plasmid; Transferase.
 FT NON TER 8
 SQ SEQUENCE 8 AA; 878 MW; EFC1B5A2CH1DD056 CRC64;

Query Match 29.5%; Score 13; DB 2; Length 8;
 Best Local Similarity 42.9%; Pred. No. 6.7e+05;
 Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 ISEMFQ 7
 DB 2 IAEVSTQ 8

RESULT 12
 P83158

ID P83158 PRELIMINARY; PRT; 8 AA.
 AC P83158;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Photosystem I iron-sulfur center (Photosystem I subunit VII) (9 kDa
 polypeptide) (PSI-C) (Fragment).
 GN Anabaena sp. (strain 131).
 OS Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
 OC NCBI_TaxID=29412;
 RN (1)
 RP SEQUENCE.
 RA Apté S.K., Uhlemann E., Schmid R., Alendort K.;
 RL Submitted (OCT-2001) to the SWISS-PROT data bank.
 CC FUNCTION: APOPROTEIN FOR THE IRON-SULFUR CENTERS PA AND FB OF THE
 CC PHOTOSYSTEM I COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
 CC 'BACTERIAL-TYPE' 4FE4S-FERRDOXIN.
 DR InterPro: IPR001450, 4FE4S-FERRDOXIN.
 DR PROSITE: PS00198, 4FE4S-FERRDOXIN; PARTIAL.
 KM Photosynthesis; Photosystem I; Iron-sulfur; 4Fe-4S; Metal-binding.
 FT NON TER 8
 SQ SEQUENCE 8 AA; 962 MW; C5B8505322D1A1F5 CRC64;

Query Match 29.5%; Score 13; DB 2; Length 8;
 Best Local Similarity 50.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 6 LQTY 9
 DB 4 VKLY 7

RESULT 13
 ID 09VRD2 PRELIMINARY; PRT; 8 AA.
 AC 09VRD2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE CG1666 Protein.
 GN CG1666.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gccayne J.D.,
 Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson R., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foeller C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howard J.J., Wei M.-H., Idegam C.,
 Jaitani W., Kalush F., Kaipen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitek R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.W., Weissbach U.,
 RA Williams S.M., Woodge T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AF03569; AAF50870.1; -;
 DR FlyBase; FBgn0040648; CG11666.
 SQ SEQUENCE 8 AA; 1062 MW; ED11B5B044004376 CRC64;

Query Match 29.5%; Score 13; DB 5; Length 8;
 Best Local Similarity 40.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 MFLQI 8
 DB 1 MWIRI 5

RESULT 14

ID Q98YK9 PRELIMINARY; PRT; 8 AA.
 AC Q98YK9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Truncated pol protein (Fragment).
 GN POL.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OK NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=991614;
 RA Schmidt B., Walter H., Moschik G., Paatz C., Werwein M., Schwingel E.,
 RA Korn K.;
 RT "Recovery of HIV-1 pol gene sequences by direct sequencing of
 RT amplification products derived from plasma samples.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF347458; AAK32535.1; -;
 FT NON TER 1
 SQ SEQUENCE 8 AA; 845 MW; 72CDB1DD736CAB8 CRC64;

Query Match 29.5%; Score 13; DB 15; Length 8;
 Best Local Similarity 40.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMF 5
 DB 4 LAEAF 8

RESULT 15

ID Q937J8 PRELIMINARY; PRT; 9 AA.
 AC Q937J8;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Putative transposition protein TniQ (Fragment).
 GN TniQ.
 OS *Escherichia coli*.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 OK NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CH210; TRANSPOSON=TN5057;
 RX MEDLINE=2160434; PubMed=11763242;
 RA Mindlin S.Z., Kholodil G.Y., Gorlenko Z.M., Minakhina S.V.,
 RA Minakhin L.S., Kalyeva E.S., Kopteva A.V., Petrova M.A.,
 RA Yurieva O.V., Nikiforov V.G.;
 RT "Mercury resistance transposons of Gram-negative environmental
 RT bacteria and their classification.";
 RL Res. Microbiol. 152:811-822(2001).
 DR EMBL; AJ302765; CAC82977.1; -;
 FT NON TER 1
 SQ SEQUENCE 9 AA; 1023 MW; 1B963AB5B7287AA4 CRC64;

Query Match 29.5%; Score 13; DB 2; Length 9;
 Best Local Similarity 40.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMF 5
 DB 5 LSDKF 9

Search completed: May 1, 2003, 23:23:33
 Job time : 30 secs

Thu May 8 16:14:13 2003

us-09-658-621b-26.ra1

Page 1

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:09:26 ; Search time 0.464286 Seconds
(without alignments)
570.332 Million cell updates/sec

Title: US-09-658-621b-26
Perfect score: 44
Sequence: 1 ISEMPLOY 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/1/1aa/5A.COMB.pdb.*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pdb.*
3: /cgn2_6/prodata/1/1aa/5C.COMB.pdb.*
4: /cgn2_6/prodata/1/1aa/5D.COMB.pdb.*
5: /cgn2_6/prodata/1/1aa/5E.COMB.pdb.*
6: /cgn2_6/prodata/1/1aa/5F.COMB.pdb.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	84.1	2035	US-08-479-537A-2	Sequence 2, Appl1
2	37	84.1	2035	US-09-083-116-2	Sequence 2, Appl1
3	37	84.1	2035	US-09-134-916A-2	Sequence 2, Appl1
4	32	72.7	2265	US-08-149-097D-36	Sequence 36, Appl1
5	32	72.7	2509	US-08-149-097D-35	Sequence 35, Appl1
6	30	68.2	272	US-09-177-165A-27	Sequence 27, Appl1
7	30	68.2	815	US-09-177-165A-24	Sequence 24, Appl1
8	29	65.9	120	US-08-727-118-2	Sequence 8, Appl1
9	29	65.9	224	US-08-248-466B-8	Sequence 8, Appl1
10	29	65.9	466	US-08-526-136-13	Sequence 13, Appl1
11	28	63.6	351	US-08-415-751-41	Sequence 41, Appl1
12	28	63.6	407	US-08-948-997-3	Sequence 3, Appl1
13	28	63.6	410	US-09-348-817A-3	Sequence 3, Appl1
14	28	63.6	462	US-08-477-451-24	Sequence 24, Appl1
15	28	63.6	487	US-08-249-112-4	Sequence 4, Appl1
16	28	63.6	487	PCT-US95-06556-4	Sequence 53, Appl1
17	28	63.6	598	US-08-853-659A-53	Sequence 53, Appl1
18	27	61.4	27	US-08-103-170-24	Sequence 24, Appl1
19	27	61.4	55	US-07-741-453A-52	Sequence 52, Appl1
20	27	61.4	65	PCT-US95-04682-7	Sequence 7, Appl1
21	27	61.4	108	US-09-114-001C-5469	Sequence 5469, Appl1
22	27	61.4	320	US-08-530-165-7	Sequence 48, Appl1
23	27	61.4	341	US-08-118-720-48	Sequence 48, Appl1
24	27	61.4	341	PCT-US93-08528-48	Sequence 48, Appl1
25	27	61.4	366	US-09-718-892-4	Sequence 4, Appl1
26	27	61.4	366	US-09-718-892-4	Sequence 4, Appl1
27	27	61.4	366	US-09-718-892-4	Sequence 4, Appl1

28	27	61.4	366	US-09-718-815-4	Sequence 4, Appl1
29	27	61.4	369	US-07-937-609-19	Sequence 19, Appl1
30	27	61.4	369	US-08-025-170-19	Sequence 19, Appl1
31	27	61.4	384	US-08-103-170-10	Sequence 10, Appl1
32	27	61.4	387	US-08-196-989B-14	Sequence 14, Appl1
33	27	61.4	387	US-08-760-936-14	Sequence 14, Appl1
34	27	61.4	451	US-08-417-330A-12	Sequence 12, Appl1
35	27	61.4	788	US-08-572-225-1	Sequence 1, Appl1
36	27	61.4	986	US-08-872-757-4	Sequence 4, Appl1
37	27	61.4	1296	US-08-728-603-15	Sequence 15, Appl1
38	27	61.4	1637	US-09-718-692-2	Sequence 2, Appl1
39	27	61.4	1637	US-09-718-852-2	Sequence 2, Appl1
40	27	61.4	1637	US-09-718-815-2	Sequence 2, Appl1
41	27	61.4	1646	US-09-535-008-67	Sequence 67, Appl1
42	27	61.4	1647	US-09-535-008-2	Sequence 2, Appl1
43	27	61.4	1649	US-09-535-008-75	Sequence 75, Appl1
44	27	61.4	1650	US-09-535-008-71	Sequence 71, Appl1
45	27	61.4	1678	US-09-535-008-69	Sequence 69, Appl1

ALIGNMENTS

RESULT 1
US-08-479-537A-2
; Sequence 2, Application US/08479537A
; Parent No. 5861381
GENERAL INFORMATION:
; APPLICANT: CHAMON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LAMH, Richard
; APPLICANT: HAPPEY, Martine
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSER: BURNS, DOANE, SWECKER & MATHEIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537A
; FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 514
; FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
; NAME: Taskin, Robin
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-6620
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537a-2
Query Match 84.1%; Score 37; DB 2; Length 2035;
Best Local Similarity 88.9%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ISEMFLOIY 9
Db 1853 ISEMFLOIY 1861
RESULT 2
US-09-083-116-2
Sequence 2, Application US/09083116
GENERAL INFORMATION:
PATENT NO. 6203795
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHY, Richard
APPLICANT: HAREVENT, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHEIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
APPLICATION DATA:
PUBLICATION NUMBER: US/09/083,116
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, AC
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CC
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2
Query Match 84.1%; Score 37; DB 4; Length 2035;
Best Local Similarity 88.9%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ISEMFLOIY 9
Db 1853 ISEMFLOIY 1861
RESULT 3
US-09-134-916A-2
Sequence 2, Application US/09134916A
PATENT NO. 6328956
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre

Thu May 8 16:14:13 2003

us-09-658-621b-26.ra1

Page 3

APPLICANT: KIENV, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HARTIVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,916A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teekin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-134-916A-2
Query Match 84.1%; Score 37; DB 4; Length 2035;
Best Local Similarity 86.9%; Pred. No. 167
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ISEMPLDQY 9
DB 1853 ISEMPLDQY 1861
RESULT 4
US-08-149-097D-36
Sequence 36 Application US/08149097D
Patent No 5874236
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: Feldman, Daniel
APPLICANT: McChae, Ann
APPLICANT: Brenner, Robert
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,097D
FILING DATE: 05-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/105,536
FILING DATE: 11-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US92/06903
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/868,354
FILING DATE: 10-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/745,206
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/620,250
FILING DATE: 30-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/482,384
FILING DATE: 20-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,751
FILING DATE: 04-APR-1989
PRIOR APPLICATION DATA:

Thu May 8 16:14:13 2003

us-09-658-621b-26 rat

Page 4

```

; APPLICATION NUMBER: WO PCT/US89/01408
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/176,899
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; TELEPHONE: (619) 238-0999
; TELEFAX: (619) 238-0062
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2265 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; FEATURE:
; OTHER INFORMATION: /product= "alpha1-2 subunit of
; US-08-149-097D-36
;
Query Match 72.7% Score 32; DB 2; Length 2265;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
OY 1 ISEMPL01Y 9
Db 530 MSSEPIKMY 538

RESULT 5
US-08-149-097D-35
; Sequence 35, Application US/08149097D
; Patent No. 5874236
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: Feldman, Daniel
; APPLICANT: McCue, Ann
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/149,097D
; FILING DATE: 05-NOV-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/105,536
; FILING DATE: 11-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US92/06903
; FILING DATE: 14-AUG-1992
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/914,231
; FILING DATE: 13-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/868,354
; FILING DATE: 10-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/745,206
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,250
; FILING DATE: 30-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/482,384
; FILING DATE: 20-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US89/01408
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/176,899
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; TELEPHONE: (619) 238-0999
; TELEFAX: (619) 238-0062
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2509 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; FEATURE:
; OTHER INFORMATION: /product= "alpha1-1 subunit of
; US-09-149-097D-35
;
Query Match 72.7% Score 32; DB 2; Length 2509;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
OY 1 ISEMPL01Y 9
Db 530 MSSEPIKMY 538

RESULT 6
US-09-177-165A-27
; Sequence 27, Application US/09177165A
; Patent No. 6426205
; GENERAL INFORMATION:
; APPLICANT: Tyers, Mike
; APPLICANT: Williams, Andrew
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
; FILE REFERENCE: 11757.10USU1
; CURRENT APPLICATION NUMBER: US/09/177,165A
; FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/092,443
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: 60/063,254
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 50
```


SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 27
LENGTH: 272
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-177-165A-27

Query Match 68.2%; Score 30; DB 4; Length 272;
Best Local Similarity 75.0%; Pred. No. 57;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 SEMFLQTY 9
DB 98 SETFLQFY 105

RESULT 7
US-09-177-165A-24
Sequence 24, Application US/09177165A
Patent No. 6426205
GENERAL INFORMATION:
APPLICANT: Tyers, Mike
APPLICANT: Williams, Andrew
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
FILE REFERENCE: 11757,10USU1
CURRENT APPLICATION NUMBER: US/09/177,165A
PRIOR APPLICATION NUMBER: 60/092,443
PRIOR FILING DATE: 1998-07-10
PRIOR APPLICATION NUMBER: 60/063,254
PRIOR FILING DATE: 1997-10-24
NUMBER OF SEQ ID NOS: 50
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 24
LENGTH: 815
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-177-165A-24

Query Match 68.2%; Score 30; DB 4; Length 815;
Best Local Similarity 75.0%; Pred. No. 196+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 SEMFLQTY 9
DB 106 SETFLQFY 113

RESULT 8
US-08-727-118-2
Sequence 2, Application US/08727118
Patent No. 5928940
GENERAL INFORMATION:
APPLICANT: MIYAZONO, KOHEI
APPLICANT: SAMBATH, KUBER T.
APPLICANT: COHEN, CHARLES M.
TITLE OF INVENTION: NOVEL MORPHOGEN-RESPONSIVE SIGNAL
TRANSDUCER AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: CREATIVE BIOMOLECULES, INC.
STREET: 45 SOUTH STREET
CITY: HOPKINTON
STATE: MA
COUNTRY: USA
ZIP: 01748
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/727,118
FILING DATE: 08-OCT-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: VITO PHD, CHRISTINE C.
REGISTRATION NUMBER: 39,061
REFERENCE/DOCKET NUMBER: CRP-121 [2054/91]
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 120 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-727-118-2

Query Match 65.9%; Score 29; DB 2; Length 120;
Best Local Similarity 55.6%; Pred. No. 39;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
DB 96 LQSEMFLQFY 104

RESULT 9
US-08-248-466B-8
Sequence 8, Application US/08248466B
Patent No. 5629182
GENERAL INFORMATION:
APPLICANT: CHOPIN, MARIE-CHRISTINE
APPLICANT: CLUZEL, PIERRE-JEAN
TITLE OF INVENTION: CLONING OF DNA FRAGMENTS ENCODING A
MECHANISM FOR RESISTANCE TO BACTERIOPHAGES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/248,466B
FILING DATE: 24-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/988,959
FILING DATE: 15-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/11381
FILING DATE: 14-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5629182man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 384-033-0 PCT FWC CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 amino acids
TYPE: amino acid
TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-248-466B-8
Query Match 65.9%; Score 29; DB 1; Length 224;
Best Local Similarity 55.6%; Pred. No. 76;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 ISEMPLOY 9
Db 143 ISEMPLOY 151

RESULT 10

US-08-526-136-13
Sequence 13, Application US/08526136
Patent No. 6107089
GENERAL INFORMATION:
APPLICANT: Towle, Christine A. et al.
TITLE OF INVENTION: ANNEXIN XI
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 507 or 558X
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526.136
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214.036
FILING DATE:
APPLICATION NUMBER: 07/837.775
FILING DATE: February 13, 1992
APPLICATION NUMBER: 07/764.465
FILING DATE: September 23, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/099001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TRIPX: 200154
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 466
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-526-136-13

Query Match 65.9%; Score 29; DB 3; Length 466;
Best Local Similarity 55.6%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ISEMPLOY 9
Db 432 IKQMFQWY 440

RESULT 11

US-08-415-751-41
Sequence 41, Application US/08415751
Patent No. 5643772
GENERAL INFORMATION:

APPLICANT: PETERSEN, CAROLYN
APPLICANT: LERCH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: PHILLIPS, MOORE, LEMPIO & FINLEY
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: California
COUNTRY: United States of America
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
COMPUTER: PC
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415.751
FILING DATE: 03-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/071.880
FILING DATE: June 1, 1993
APPLICATION NUMBER: 07/841.301
FILING DATE: May 29, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Hana Dolazalova
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-1677
TELEFAX: (415) 324-1678
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 351 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Cryptosporidium parvum
FEATURE:
NAME/KEY: Positions coded by nonsense codons are
NAME/KEY: Identified as Xaa.

US-08-415-751-41

Query Match 63.6%; Score 28; DB 1; Length 351;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMPLOY 9
Db 7 IAEVLQPY 15

RESULT 12

US-08-948-997-3
Sequence 3, Application US/08948997
Patent No. 6008020
GENERAL INFORMATION:
APPLICANT: HASTINGS, GREGG
APPLICANT: COLEMAN, TIM
APPLICANT: LAWRENCE, DANIEL
TITLE OF INVENTION: BRAIN-ASSOCIATED INHIBITOR OF
TITLE OF INVENTION: TISSUE-TYPE PLASMINOGEN ACTIVATOR
NUMBER OF SEQUENCES: 17

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
;; STREET: 9410 KEY WEST AVENUE
;; CITY: ROCKVILLE
;; STATE: MD
;; COUNTRY: USA
;; ZIP: 20850
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: FLOPPY disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/948,997
;; FILING DATE: Oct-10-97
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: A. ANDERS BROOKES
;; REGISTRATION NUMBER: 36,373
;; REFERENCE/DOCKET NUMBER: PP336
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (301) 309-8504
;; TELEFAX: (301) 309-8512
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 407 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-948-997-3

Query Match 63.6%; Score 28; DB 3; Length 407;
Best Local Similarity 62.5%; Pred. No. 2.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 8
DB 118 VSEKFLQI 125

RESULT 13
US-09-348-817A-3
; Sequence 3, Application US/09348817A
; Patent No. 6191260
; GENERAL INFORMATION:
; APPLICANT: Hastings et al.
; TITLE OF INVENTION: Brain-Associated Inhibitor of Tissue-Type Plasminogen
; TITLE OF INVENTION: Activator
; FILE REFERENCE: PP336D1
; CURRENT APPLICATION NUMBER: US/09/348,817A
; CURRENT FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 08/948,997
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/028,117
; PRIOR FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Gallus gallus
US-09-348-817A-3

Query Match 63.6%; Score 28; DB 4; Length 410;
Best Local Similarity 62.5%; Pred. No. 2.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 8
DB 120 VSEKFLQI 127

RESULT 14
US-08-477-451-24
; Sequence 24, Application US/08477451
; Patent No. 592865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335,002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;; US-08-477-451-24

Query Match 63.6%; Score 28; DB 2; Length 462;
Best Local Similarity 57.1%; Pred. No. 2.7e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 EMFLQIY 9
DB 229 QIFLOIY 235

RESULT 15
US-08-249-112-4
; Sequence 4, Application US/08249112
; Patent No. 5527703
; GENERAL INFORMATION:
; APPLICANT: Cully, Doris F.
; APPLICANT: Arena, Joseph P.
; APPLICANT: Liu, Ken K.
; APPLICANT: Vassiliadis, Demetrios
; TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE
; TITLE OF INVENTION: CHANNELS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Waller, John W.
; STREET: 126 E. Lincoln Ave., P.O. Box 2000
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/249,112
 ; FILING DATE: 25-MAY-1994
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wallen, John W.
 ; REGISTRATION NUMBER: 35,403
 ; REFERENCE/DOCKET NUMBER: 19194
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (908) 594-3905
 ; TELEFAX: (908) 594-4720
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 487 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-249-112-4

Query Match 63.6%; Score 28; DB 1; Length 487;
 Best Local Similarity 55.6%; Pred. No. 2.8e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ISEMPLQIY 9
 : ||: ||
 Db 140 MEMPLRIY 148

Search completed: May 1, 2003, 23:14:27
 Job time : 1.46429 secs

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 22:59:30 ; Search time 0.321429 Seconds
(without alignments)

1161.337 Million cell updates/sec

Title: US-09-658-621B-26

Perfect score: 44
Sequence: 1 ISEMFQIY 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	1255	MUC1_HUMAN	P15941 h mucin 1 P
2	35	79.5	475	MUC1_HYLLA	Q29435 hyllobates l
3	33	75.0	2223	CCAE_DISOM	P56699 discopryge o
4	32	72.7	1848	CCAA_DROME	P91645 drosophila
5	32	72.7	2164	CCAA_MOUSE	P97445 mus musculu
6	32	72.7	2212	CCAA_RAT	P54282 rattus norv
7	32	72.7	2424	CCAA_RABIT	P27884 coryctolagus
8	32	72.7	2505	CCAA_HUMAN	O00555 homo sapien
9	31	70.5	144	SODM_PALVU	P28765 palinurus v
10	31	70.5	366	VGLM_HSVSA	O01017 herpesvirus
11	31	70.5	491	VOR4_MYCTU	O53209 mycobacteri
12	31	70.5	604	VEJA_ECOLI	P33913 escherichia
13	31	70.5	896	APCE_SYNY4	Q02907 synechocyst
14	31	70.5	1071	PR16_YEAST	P15938 saccharomyc
15	31	70.5	1630	MSPI_PLAFK	P04933 plasmodium
16	31	70.5	1639	MSPI_PLAFW	Q10105 schizosach
17	31	70.5	2670	YAOS_SCHPO	Q9ycs9 dirosophila
18	30	68.2	387	O94A_DROME	O02496 mus musculu
19	30	68.2	630	MUC1_MOUSE	P35939 saccharomyc
20	30	68.2	772	PR1E_YEAST	Q12018 saccharomyc
21	30	68.2	815	CC53_YEAST	Q12751 saccharomyc
22	30	68.2	981	YV48_YEAST	Q12190 porcine rot
23	30	68.2	1082	RROB_RORPC	Q91106 nephroselm
24	30	68.2	1109	RROC_NEPOL	O92939 chlamydia p
25	30	68.2	1393	RROC_CHLPM	O92939 chlamydia p
26	30	68.2	1396	RROC_CHLPM	O92939 chlamydia p
27	30	68.2	1396	RROC_CHLPM	O92939 chlamydia p
28	29	65.9	348	HEP2_BACST	P31114 bacillus su
29	29	65.9	463	ANK7_MOUSE	O07076 mus musculu
30	29	65.9	466	ANK7_HUMAN	P20073 mus musculu
31	29	65.9	527	SYK_CHLPM	O92939 chlamydia p
32	29	65.9	548	ALGB_YEAST	P53954 saccharomyc
33	29	65.9	558	ATVA_SYNY3	P73866 synechocyst

34	29	65.9	558	YNT3_YEAST	P53870 saccharomyc
35	29	65.9	598	LEPA_MYCPN	P75498 mycoplasma
36	29	65.9	673	UYRB_BORBU	O51776 borrelia bu
37	29	65.9	806	SECA_MYCPN	P47318 mycoplasma
38	29	65.9	808	SECA_MYCPN	P75559 mycoplasma
39	29	65.9	1058	U202_ARATH	O92144 arabidopsis
40	29	65.9	1071	TR1_THESC	P66086 thermoplas
41	29	65.9	1122	RREP_P1ZHT	P26676 human para
42	28	63.6	112	Y122_CABEL	P03935 caenorhabdi
43	28	63.6	117	Y627_METJA	O58044 methanococ
44	28	63.6	188	Y101_UREPA	O9pr43 ureaplasma
45	28	63.6	299	YX06_CABEL	Q11113 caenorhabdi

ALIGNMENTS

RESULT 1
MUC1_HUMAN STANDARD; PRT; 1255 AA.
AC P15941 P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4U2;
AC Q16615; Q14876; Q9UE75; Q9UE76; Q9UOL1; Q9BXA4;
AC Q1-UAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)
DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)
DE (Tumor-associated epithelial membrane antigen) (EMA) (H23AG) (Peanut-
DE reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen
DE DF3) (CD227 antigen).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Pancereas;
RX MEDLINE=90368716; PubMed=2394722;
RA Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A AND B).
RX MEDLINE=9020794; PubMed=2318825;
RA Legendre M.-J., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Episialin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini.";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM SEC).
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;
RA Wreschner D.H., Hareuveni M., Tsarfay I., Smorodinsky N., Horev J.,
RA Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.;

RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Harevuani M., Tsarfaty I., Zaretsky J., Kocles P., Horev J.,
RA Zrihan S., Weiss M., Green S., Lache R., Keydar I., Wreschner D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transcribed gene and over-expression in breast cancer
RT tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=91033045; PubMed=1688329;
RA Tsarfaty I., Harevuani M., Horev J., Zaretsky J., Weiss M.,
RA Jeltsch J.M., Garnier J.M., Lache R., Keydar I., Wreschner D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RL Gene 93:313-318(1990).
RN [8]
RP SEQUENCE FROM N.A. (ISOFORM Y).
RX MEDLINE=95010060; PubMed=7925397;
RA Zrihan-Licht S., Vos H.L., Barnuch A., Elroy-Stein O., Sagiv D.,
RA Keydar I., Hilkens J., Wreschner D.H.;
RT "Characterization and molecular cloning of a novel MUC1 protein,
RT devoid of tandem repeats, expressed in human breast cancer tissue.";
RL Eur. J. Biochem. 224:787-795(1994).
RN [9]
RP SEQUENCE FROM N.A. (ISOFORMS X; Y AND Z).
RX MEDLINE=97355747; PubMed=9212228;
RA Oosterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,
RA Finstad C.L.;
RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
RT cancer cell lines and demonstration of a new short variant form
RT (MUC-1/Z).";
RL Int. J. Cancer 72:87-94(1997).
RN [10]
RP SEQUENCE FROM N.A. (ISOFORM Y).
RX Zhang L.X., Li C.H.;
RT "Molecular cloning of an isoform of MUC1, MUC1/Y.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A. (ISOFORM S).
RC TISSUE=Epithelial cancer;
RX Zhang L.X., Li C.H., Sun L.Y., Yue W.;
RT "Cloning of a new potential secreted short variant form of MUC1 mucin
RT in epithelial cancer cell line.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [12]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88330762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [13]
RP SEQUENCE OF 1-160 FROM N.A. (ISOFORM B).
RX MEDLINE=90086473; PubMed=2597151;
RA Abe M., Siddiqui J., Kufe D.;
RT "Sequence analysis of the 5' region of the human DF3 breast
RT carcinoma-associated antigen gene.";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [14]
RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM B).
RC TISSUE=Thyroid;
RX MEDLINE=96183746; PubMed=860866;
RA Weiss M., Barnuch A., Keydar I., Wreschner D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene.";
RN [15]
RL Int. J. Cancer 66:55-59(1996).
RN [15]
RP SEQUENCE OF 1-89 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA Lee L.N., Luh K.T., Wu C.W.;
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT tissues.";
RL Oncology 53:118-126(1996).
RN [16]
RP SEQUENCE OF 1-46 FROM N.A. (ISOFORMS C AND D).
RC TISSUE=Breast carcinoma;
RA Baluwela L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
RN [17]
RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX MEDLINE=97460054; PubMed=9312074;
RA Mueller S., Goletz S., Packer N., Goolley A.A., Lawson A.M.,
RA Hantsch F.-G.;
RT "Localization of O-glycosylation sites on glycopeptide fragments from
RT lactation-associated MUC1. All putative sites within the tandem
RT repeat are glycosylation targets in vivo.";
RL J. Biol. Chem. 272:24780-24793(1997).
RN [18]
RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX MEDLINE=99303572; PubMed=10373415;
RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Goolley A.A.,
RA Hantsch F.-G.;
RT "High density O-glycosylation on tandem repeat peptide from secretory
RT MUC1 of T47D breast cancer cells.";
RL J. Biol. Chem. 274:18165-18172(1999).
RN [19]
RP POLYMORPHISM WITHIN THE REPEAT.
RX MEDLINE=21359366; PubMed=11350974;
RA Engelmann K., Baldus S.B., Hantsch F.-G.;
RT "Identification and topology of variant sequences within individual
RT repeat domains of the human epithelial tumor mucin MUC1.";
RL J. Biol. Chem. 276:27764-27769(2001).
RN [20]
RP CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
RX MEDLINE=99211485; PubMed=10197628;
RA Barnuch A., Hartmann M.-L., Yoeli M., Adereeth Y., Greenstein S.,
RA Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,
RA Wreschner D.H.;
RT "The breast cancer-associated MUC1 gene generates both a receptor and
RT its cognate binding protein.";
RL Cancer Res. 59:1552-1561(1999).
RN [21]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.
RX MEDLINE=21240104; PubMed=11341784;
RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
RA Harris A.;
RT "Identification of MUC1 proteolytic cleavage sites in vivo.";
RL Biochem. Biophys. Res. Commun. 283:715-720(2001).
RN [22]
RP CHARACTERIZATION.
RX MEDLINE=21835452; PubMed=11847293;
RA Wreschner D.H., McGuckin M.A., Williams S.J., Barnuch A., Yoeli M.,
RA Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,
RA Stacey M., Lin H.-H., Gordon S.;
RT "Generation of ligand-receptor alliances by 'SEA' module-mediated
RT cleavage of membrane-associated mucin proteins.";
RL Protein Sci. 11:698-706(2002).
RN [23]
RP PHOSPHORYLATION.
RX MEDLINE=95080414; PubMed=7988707;
RA Zrihan-Licht S., Barnuch A., Elroy-Stein O., Keydar I., Wreschner D.H.;
RT "Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins.
RT Cytokine receptor-like molecules.";
RL FEBS Lett. 356:130-136(1994).
RN [24]
CC -I- FUNCTION: May play a role in adhesive functions and in cell-cell
interactions, metastasis and signaling. May provide a protective

Query Match 100.0%; Score 44; DB 1; Length 1255;
 Best Local Similarity Pred. No. 0.14;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
 DB 1073 ISEMFLQTY 1081

RESULT 2
 ID MUC1_HYLLA STANDARD; PRT; 475 AA.
 AC Q29435;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Mucin 1 precursor (MOC-1).
 GN MUC1.
 OS Hylobates lat (Common gibbon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
 OX NCBI_TaxID=9580;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96351712; PubMed=8747930;
 RA Spicer A.P., Duhig T., Chilton B.S., Gendler S.J.;
 RT "Analysis of mammalian MUC1 genes reveals potential functionally
 important domains."
 RL Mamm. Genome 6:885-888(1995).
 CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
 CYTOSKELETON (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: IMA589; AAA69965.1; -;
 DR EMBL: LA1625; AAA69918.1; -;
 DR EMBL: LA1624; AAA69918.1; JOINED.
 DR InterPro: IPR000082; SEA_domain.
 DR Pfam: PF01390; SEA; 1.
 DR SMART: SM00200; SEA; 1.
 DR PROSITE: PS50024; SEA; 1.
 KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
 Repeat.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 475 MUCIN 1.
 FT DOMAIN 24 380 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 381 401 POTENTIAL.
 FT DOMAIN 402 475 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 102 121 1.
 FT REPEAT 122 141 2.
 FT REPEAT 142 161 3.
 FT REPEAT 162 181 4.
 FT DOMAIN 254 371 SEA.
 SQ SEQUENCE 475 AA; 49371 MW; D7A699D6D68C6622 CRC64;

Query Match 79.5%; Score 35; DB 1; Length 475;
 Best Local Similarity Pred. No. 4.5;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
 DB 293 ISEMFLQTY 301

RESULT 3
 ID CCAE_DISOM STANDARD; PRT; 2223 AA.
 AC p56693;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable voltage-dependent R-type calcium channel alpha-1E subunit
 (DOE-1).
 OS Discosoma ommata (Electric ray).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squalia; Hyngosqualea; Pristiogalea; Batoidae;
 OC Torpediniformes; Narcinidae; Narcinidae; Discosoma.
 OX NCBI_TaxID=7785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Electric lobe;
 RX MEDLINE=93248175; PubMed=7683405;
 RA Horne W.A., Ellinger P.T., Imman I., Zhou M., Tsien R.W., Schwarz T.L.;
 RT "Molecular diversity of Ca2+ channel alpha 1 subunits from the marine
 ray Discosoma ommata."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:3787-3791(1993).
 CC -1- FUNCTION: THE ISOFORM ALPHA-1E GIVES RISE TO THE "HIGH-VOLTAGE
 CURRENTS. R-TYPE CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE
 ACTIVATED" (HVA) GROUP (BY SIMILARITY).
 CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
 COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
 IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
 FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
 SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
 CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
 LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY (BY
 SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSION IS HIGHER IN THE FOREBRAIN THAN IN
 THE ELECTRIC LOBE.
 CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
 HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
 POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
 PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
 SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
 CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
 FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L12531; -; NOT ANNOTATED CDS.
 DR InterPro: IPR001682; Ca/Na_pore.
 DR InterPro: IPR002077; Ca channel.
 DR InterPro: IPR002111; Ca channel_Tmpl.
 DR InterPro: IPR000636; M-channel_mtg.
 DR Pfam: PF00520; Ion trans; 4.
 DR PRINTS: PRO0167; CACCHANNE.
 KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 KW Calcium channel; Glycoprotein; Repeat; Multigene family;
 KW Calcium-binding; Phosphorylation.
 FT REPEAT 89 366 I.
 FT REPEAT 478 720 II.
 FT REPEAT 1092 1377 III.
 FT REPEAT 1414 1666 IV.
 FT DOMAIN 1 102 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 103 121 S1 OF REPEAT 1 (POTENTIAL).
 FT DOMAIN 122 139 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 140 159 S2 OF REPEAT 1 (POTENTIAL).
 FT DOMAIN 160 171 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 172 189 S3 OF REPEAT 1 (POTENTIAL).

FT	DOMAIN	190	194	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	195	213	S4 OF REPEAT I (POTENTIAL).
FT	DOMAIN	214	232	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	233	252	S5 OF REPEAT I (POTENTIAL).
FT	DOMAIN	253	338	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	339	363	S6 OF REPEAT I (POTENTIAL).
FT	DOMAIN	364	490	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	491	509	S1 OF REPEAT II (POTENTIAL).
FT	DOMAIN	510	524	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	525	544	S2 OF REPEAT II (POTENTIAL).
FT	DOMAIN	545	552	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	553	571	S3 OF REPEAT II (POTENTIAL).
FT	DOMAIN	572	581	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	582	600	S4 OF REPEAT II (POTENTIAL).
FT	DOMAIN	601	619	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	620	639	S5 OF REPEAT II (POTENTIAL).
FT	DOMAIN	640	692	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	693	717	S6 OF REPEAT II (POTENTIAL).
FT	DOMAIN	718	1105	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1106	1124	S1 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1125	1140	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1141	1160	S2 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1161	1172	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1173	1191	S3 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1192	1205	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1206	1224	S4 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1225	1243	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1244	1263	S5 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1264	1349	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1350	1374	S6 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1375	1429	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1430	1448	S1 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1449	1463	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1464	1483	S2 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1484	1491	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1492	1510	S3 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1511	1519	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1520	1538	S4 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1539	1557	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1558	1577	S5 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1578	1638	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1639	1664	S6 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1664	2223	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2223	33	POLY-ALA.
FT	DOMAIN	33	40	POLY-SER.
FT	TRANSMEM	40	735	POLY-SER.
FT	DOMAIN	735	1068	BINDING TO THE BETA SUBUNIT (BY
FT	TRANSMEM	1064	403	SIMILARITY).
FT	DOMAIN	386	321	CALCIUM ION SELECTIVITY AND PERMEABILITY
FT	TRANSMEM	321	321	(BY SIMILARITY).
FT	DOMAIN	671	671	CALCIUM ION SELECTIVITY AND PERMEABILITY
FT	TRANSMEM	671	671	(BY SIMILARITY).
FT	DOMAIN	1323	1323	CALCIUM ION SELECTIVITY AND PERMEABILITY
FT	TRANSMEM	1323	1323	(BY SIMILARITY).
FT	DOMAIN	1611	1611	CALCIUM ION SELECTIVITY AND PERMEABILITY
FT	TRANSMEM	1611	1611	(BY SIMILARITY).
FT	DOMAIN	439	450	PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT	TRANSMEM	1674	1674	BY SIMILARITY.
FT	DOMAIN	1692	1703	PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT	TRANSMEM	287	287	BY SIMILARITY.
FT	DOMAIN	1300	1300	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	TRANSMEM	1300	1300	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	DOMAIN	1519	1519	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	TRANSMEM	1519	1519	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	DOMAIN	2223	2223	MM: 59722DC03B1CFC5B CXC64;
FT	TRANSMEM	2223	2223	MM: 59722DC03B1CFC5B CXC64;

Query Match 75.0%; Score 33; DB 1; Length 2223;
 Best Local Similarity 75.0%; Pred. No. 65; Mismatches 2; Indels 0; Gaps 0;

Db 535 SEMFLXMY 542

RESULT 4	CCAA_DROME	STANDARD:	PRT, 1848 AA.
ID	CCAA_DROME		
AC	P91645: C011713; C011714; Q9VYR8;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DE	Voltage-dependent calcium channel type A alpha-1 subunit (Cacophony protein) (Nightblind A protein) (No-on-transient B protein) (DMCA1A).		
DE	CAC OR NBA OR NONB OR CG1522.		
GN	Drosophila melanogaster (Fruit fly).		
OS	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;		
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxId=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=97141514; PubMed=8987815;		
RX	Smith L.A., Wang X.J., Peixoto A.A., Neumann E.K., Hall L.M.,		
RA	Hall J.C.;		
RT	"A Drosophila calcium channel alpha subunit gene maps to a genetic		
RT	locus associated with behavioral and visual defects."		
RL	J. Neurosci. 16:7868-7879 (1996).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Berkeley;		
RX	MEDLINE=20196006; PubMed=10731132;		
RX	Adams W.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,		
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abri J.F., Achyari A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Betman B.P., Bhandal D., Borkstein P.,		
RA	Borison D., Botchan M.R., Bouck J., Brokstein P., Brotshakov S.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K.E., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,		
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honck J.,		
RA	Hoskins R., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,		
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.R., Ketchum K.A.,		
RA	Kamel B.E., Kodira C.D., Krat C., Kravitz S., Kulp D., Lai Z.,		
RA	Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson X.A., Nixon K., Nuskern D.R., Paoleb J.M.,		
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		
RA	Reichert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,		
RA	Shue B.C., Sideri-Kiamos I., Simpson M., Skupski M.P., Smith T.,		
RA	Spiel E., Spradling A.C., Stapleton M., Strong R., Sun E.,		
RA	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,		
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,		
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,		
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,		
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;		
RT	"The genome sequence of Drosophila melanogaster."		
RL	Science 287:2185-2195 (2000).		
RP	[3]		
RP	PARTIAL SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.		
RX	MEDLINE=97247730; PubMed=903853;		
RA	Peixoto A.A., Smith L.A., Hall J.C.;		
RT	"Genomic organization and evolution of alternative exons in a		
RT	Drosophila calcium channel gene."		
RL	Genetics 145:1003-1013 (1997).		
RN	[4]		

RP VARIANT CAC-S.
 RX MEDLINE=9835104; PubMed=9649530;
 RA Smith L.A., Peixoto A.A., Kramer E.M., Villella A., Hall J.C.;
 RT "Courtship and visual defects of cacophony mutants impair functional
 complexity of a calcium-channel alpha subunit in *Drosophila*.";
 RL Genetics 149:1407-1426(1998).
 CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIANE THE
 ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
 IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
 CONTRACTION, NEUROTRANSMITTER RELEASE, GENE EXPRESSION, CELL
 MOTILITY, CELL DIVISION AND CELL DEATH (BY SIMILARITY). PROBABLY
 ENCODES A DIHYDROPYRIDINE-INSENSITIVE CURRENT. VITAL FOR SURVIVAL
 TO ADULTHOOD.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: A number of isoforms are produced by
 alternative splicing. At least 2 regions (184 and 1/11) undergo
 alternative splicing. The total number of isoforms is currently
 not known.
 CC -1- TISSUE SPECIFICITY: EXPRESSED WIDELY IN THE EMBRYONIC NERVOUS
 SYSTEM.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION PEAKS IN THE FIRST LARVAL INSTAR,
 MIDPUPAL, AND LATE PUPAL STAGES. IN LATE-STAGE EMBRYOS, IT IS
 EXPRESSED PREFERENTIALLY IN THE NERVOUS SYSTEM.
 CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
 HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
 POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
 PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
 SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
 CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNIT
 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, U55776; AAC7406.1; -;
 DR EMBL, AE003487; AAF48120.1; -;
 DR EMBL, U88664; AAB53271.1; -;
 DR EMBL, U88665; AAB53272.1; -;
 DR FLYBASE; FBGN0005563; cac.
 DR InterPro; IPR001682; Ca/Na pore.
 DR InterPro; IPR002077; Ca channel.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000636; M-channel_nlg.
 DR InterPro; IPR003915; PKD_2.
 DR Pfam; PF00036; efhand; 2.
 DR Pfam; PF00520; ion_trans; 4.
 DR PRINTS; PRO0167; CACCHANNEL.
 DR PRINTS; PRO1433; POLYCYSTIN2.
 DR SMART; SM00054; Efh; 1.
 KM Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 KM Calcium channel; Glycoprotein; Repeat; Multigene family;
 KM Calcium-binding; Phosphorylation; Alternative splicing;
 KM Developmental protein.
 FT REPEAT 25 316
 FT REPEAT 316 670
 FT REPEAT 427 1049
 FT REPEAT 762 1344
 FT REPEAT 1086 1344
 FT DOMAIN 1 38
 FT TRANSMEM 39 57
 FT DOMAIN 58 75
 FT TRANSMEM 76 95
 FT DOMAIN 96 107
 FT TRANSMEM 108 128
 FT DOMAIN 129 133
 FT TRANSMEM 134 152
 FT DOMAIN 153 171
 FT TRANSMEM 172 191

FT DOMAIN 192 288
 FT TRANSMEM 289 313
 FT DOMAIN 314 441
 FT TRANSMEM 442 460
 FT DOMAIN 461 475
 FT TRANSMEM 476 495
 FT DOMAIN 496 503
 FT TRANSMEM 504 522
 FT DOMAIN 523 531
 FT TRANSMEM 532 550
 FT DOMAIN 551 569
 FT TRANSMEM 570 589
 FT DOMAIN 590 642
 FT TRANSMEM 643 667
 FT DOMAIN 668 767
 FT TRANSMEM 768 786
 FT DOMAIN 787 802
 FT TRANSMEM 803 822
 FT DOMAIN 823 834
 FT TRANSMEM 835 853
 FT DOMAIN 854 866
 FT TRANSMEM 867 885
 FT DOMAIN 886 904
 FT TRANSMEM 905 924
 FT DOMAIN 925 1013
 FT TRANSMEM 1014 1038
 FT DOMAIN 1039 1093
 FT TRANSMEM 1094 1122
 FT DOMAIN 1123 1127
 FT TRANSMEM 1128 1147
 FT DOMAIN 1148 1155
 FT TRANSMEM 1156 1174
 FT DOMAIN 1175 1181
 FT TRANSMEM 1182 1200
 FT DOMAIN 1201 1219
 FT TRANSMEM 1220 1239
 FT DOMAIN 1240 1305
 FT TRANSMEM 1306 1330
 FT DOMAIN 1331 1348
 FT TRANSMEM 1349 1367
 FT DOMAIN 1368 1374
 FT TRANSMEM 1374 1374
 FT CARBOHYD 234 234
 FT CARBOHYD 235 235
 FT CARBOHYD 865 865
 FT VARSPLIC 121 153
 FT FT
 FT VARSPLIC 315 352
 FT FT
 Query Match 72.7%
 Best Local Similarity 55.6%
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ISEMFLOY 9
 Db 485 NSEMFKEY 493

RESULT 5
 ID CCA MOUSE STANDARD; PRT; 2164 AA.
 AC P97445
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Voltage-dependent P/Q-type calcium channel alpha-1A subunit (Calcium
DE channel, L type, alpha-1 polypeptide isoform 4) (Brain calcium
DE channel 1) (B1).
GN CACNA1A OR CACNA1A4 OR CCH4A OR CACNA4 OR CACNA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT TG LEU-601.
RC STRAIN=DSB/2J.
RA MEDLINE=97083572; PubMed=8929530;
RA Fletcher C.F., Lutz C.M., O'Sullivan T.N., Shaughnessy J.D. Jr.,
RA Hawkes R., Frankel W.N., Copeland N.G., Jenkins N.A.;
RT "Absence epilepsy in tottering mutant mice is associated with calcium
RT channel defects".
RL Cell 87:607-617(1996).
CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDATE THE
CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC IN A VARIETY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC IN A VARIETY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
CC GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/Q-TYPE CALCIUM
CC CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
CC ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-
CC IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO
CC DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-
CC GVIA).
CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC; MAINLY FOUND IN THE
CC CEREBELLUM, OLFACTORY BULB, CEREBRAL CORTEX, HIPPOCAMPUS, AND
CC INFERIOR COLICULUS. IN THE HIPPOCAMPUS, EXPRESSION OCCURS IN
CC PYRAMIDAL AND GRANULE NEURONS, AS WELL AS IN INTERNEURONS.
CC PURKINJE CELLS CONTAIN PREDOMINANTLY P-TYPE VSCC. THE Q-TYPE BEING
CC A PROMINENT CALCIUM CURRENT IN CEREBELLAR GRANULE CELLS.
CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -1- DISEASE: DEFECTS IN CACNA1A ARE THE CAUSE OF A DELAYED-ONSET,
CC RECESSIVE NEUROLOGICAL DISORDER SEEN IN TOTTERING (TG) MUTANTS,
CC RESULTING IN ATAXIA, MOTOR SEIZURES AND BEHAVIORAL ABSENCE
CC SEIZURES RESEMBLING PETIT MAL EPILEPSY (OR ABSENCE EPILEPSY) IN
CC HUMANS. THERE ARE TWO MORE ALLELES: LEANER (TG(LA)), THAT IS
CC CHARACTERIZED BY SEVERE ATAXIA AND FREQUENT DEATH PAST MEANING,
CC BUT NO MOTOR SEIZURES; AND ROLLING (NGOTA (TG(ROL))), THAT PRESENTS
CC AN INTERMEDIARY PHENOTYPE. THE ATAXIA BEING SOMEWHAT MORE SEVERE
CC THAT WITH TG, BUT WITHOUT MOTORS SEIZURES. SELECTIVE DEGENERATION
CC OF CEREBELLAR PURKINJE CELLS HAS BEEN SHOWN FOR ALL THESE TYPES OF
CC MUTANTS.
CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U76716; AAC52940.1; -
CC MGD: MG1:109482; CACNA1A.
CC InterPro: IPR000637; AT hook.
CC InterPro: IPR001692; Ca/Na_pore.

DR InterPro: IPR002077; Ca_channel.
DR InterPro: IPR002111; Cat_channel_trpL.
DR InterPro: IPR000636; M+channel_nig.
DR Pfam: PF00520; Ion_trans_4.
DR PRINTS: PR00167; CACCHANNEL.
DR SMART: SM00384; AT_hook_1.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel; Glycoprotein; Repeat; Multigene family;
KW Calcium-binding; Phosphorylation; Disease mutation.
FT REPEAT 1
FT REPEAT 2
FT REPEAT 3
FT REPEAT 4
FT REPEAT 5
FT REPEAT 6
FT REPEAT 7
FT REPEAT 8
FT REPEAT 9
FT REPEAT 10
FT REPEAT 11
FT REPEAT 12
FT REPEAT 13
FT REPEAT 14
FT REPEAT 15
FT REPEAT 16
FT REPEAT 17
FT REPEAT 18
FT REPEAT 19
FT REPEAT 20
FT REPEAT 21
FT REPEAT 22
FT REPEAT 23
FT REPEAT 24
FT REPEAT 25
FT REPEAT 26
FT REPEAT 27
FT REPEAT 28
FT REPEAT 29
FT REPEAT 30
FT REPEAT 31
FT REPEAT 32
FT REPEAT 33
FT REPEAT 34
FT REPEAT 35
FT REPEAT 36
FT REPEAT 37
FT REPEAT 38
FT REPEAT 39
FT REPEAT 40
FT REPEAT 41
FT REPEAT 42
FT REPEAT 43
FT REPEAT 44
FT REPEAT 45
FT REPEAT 46
FT REPEAT 47
FT REPEAT 48
FT REPEAT 49
FT REPEAT 50
FT REPEAT 51
FT REPEAT 52
FT REPEAT 53
FT REPEAT 54
FT REPEAT 55
FT REPEAT 56
FT REPEAT 57
FT REPEAT 58
FT REPEAT 59
FT REPEAT 60
FT REPEAT 61
FT REPEAT 62
FT REPEAT 63
FT REPEAT 64
FT REPEAT 65
FT REPEAT 66
FT REPEAT 67
FT REPEAT 68
FT REPEAT 69
FT REPEAT 70
FT REPEAT 71
FT REPEAT 72
FT REPEAT 73
FT REPEAT 74
FT REPEAT 75
FT REPEAT 76
FT REPEAT 77
FT REPEAT 78
FT REPEAT 79
FT REPEAT 80
FT REPEAT 81
FT REPEAT 82
FT REPEAT 83
FT REPEAT 84
FT REPEAT 85
FT REPEAT 86
FT REPEAT 87
FT REPEAT 88
FT REPEAT 89
FT REPEAT 90
FT REPEAT 91
FT REPEAT 92
FT REPEAT 93
FT REPEAT 94
FT REPEAT 95
FT REPEAT 96
FT REPEAT 97
FT REPEAT 98
FT REPEAT 99
FT REPEAT 100
FT REPEAT 101
FT REPEAT 102
FT REPEAT 103
FT REPEAT 104
FT REPEAT 105
FT REPEAT 106
FT REPEAT 107
FT REPEAT 108
FT REPEAT 109
FT REPEAT 110
FT REPEAT 111
FT REPEAT 112
FT REPEAT 113
FT REPEAT 114
FT REPEAT 115
FT REPEAT 116
FT REPEAT 117
FT REPEAT 118
FT REPEAT 119
FT REPEAT 120
FT REPEAT 121
FT REPEAT 122
FT REPEAT 123
FT REPEAT 124
FT REPEAT 125
FT REPEAT 126
FT REPEAT 127
FT REPEAT 128
FT REPEAT 129
FT REPEAT 130
FT REPEAT 131
FT REPEAT 132
FT REPEAT 133
FT REPEAT 134
FT REPEAT 135
FT REPEAT 136
FT REPEAT 137
FT REPEAT 138
FT REPEAT 139
FT REPEAT 140
FT REPEAT 141
FT REPEAT 142
FT REPEAT 143
FT REPEAT 144
FT REPEAT 145
FT REPEAT 146
FT REPEAT 147
FT REPEAT 148
FT REPEAT 149
FT REPEAT 150
FT REPEAT 151
FT REPEAT 152
FT REPEAT 153
FT REPEAT 154
FT REPEAT 155
FT REPEAT 156
FT REPEAT 157
FT REPEAT 158
FT REPEAT 159
FT REPEAT 160
FT REPEAT 161
FT REPEAT 162
FT REPEAT 163
FT REPEAT 164
FT REPEAT 165
FT REPEAT 166
FT REPEAT 167
FT REPEAT 168
FT REPEAT 169
FT REPEAT 170
FT REPEAT 171
FT REPEAT 172
FT REPEAT 173
FT REPEAT 174
FT REPEAT 175
FT REPEAT 176
FT REPEAT 177
FT REPEAT 178
FT REPEAT 179
FT REPEAT 180
FT REPEAT 181
FT REPEAT 182
FT REPEAT 183
FT REPEAT 184
FT REPEAT 185
FT REPEAT 186
FT REPEAT 187
FT REPEAT 188
FT REPEAT 189
FT REPEAT 190
FT REPEAT 191
FT REPEAT 192
FT REPEAT 193
FT REPEAT 194
FT REPEAT 195
FT REPEAT 196
FT REPEAT 197
FT REPEAT 198
FT REPEAT 199
FT REPEAT 200
FT REPEAT 201
FT REPEAT 202
FT REPEAT 203
FT REPEAT 204
FT REPEAT 205
FT REPEAT 206
FT REPEAT 207
FT REPEAT 208
FT REPEAT 209
FT REPEAT 210
FT REPEAT 211
FT REPEAT 212
FT REPEAT 213
FT REPEAT 214
FT REPEAT 215
FT REPEAT 216
FT REPEAT 217
FT REPEAT 218
FT REPEAT 219
FT REPEAT 220
FT REPEAT 221
FT REPEAT 222
FT REPEAT 223
FT REPEAT 224
FT REPEAT 225
FT REPEAT 226
FT REPEAT 227
FT REPEAT 228
FT REPEAT 229
FT REPEAT 230
FT REPEAT 231
FT REPEAT 232
FT REPEAT 233
FT REPEAT 234
FT REPEAT 235
FT REPEAT 236
FT REPEAT 237
FT REPEAT 238
FT REPEAT 239
FT REPEAT 240
FT REPEAT 241
FT REPEAT 242
FT REPEAT 243
FT REPEAT 244
FT REPEAT 245
FT REPEAT 246
FT REPEAT 247
FT REPEAT 248
FT REPEAT 249
FT REPEAT 250
FT REPEAT 251
FT REPEAT 252
FT REPEAT 253
FT REPEAT 254
FT REPEAT 255
FT REPEAT 256
FT REPEAT 257
FT REPEAT 258
FT REPEAT 259
FT REPEAT 260
FT REPEAT 261
FT REPEAT 262
FT REPEAT 263
FT REPEAT 264
FT REPEAT 265
FT REPEAT 266
FT REPEAT 267
FT REPEAT 268
FT REPEAT 269
FT REPEAT 270
FT REPEAT 271
FT REPEAT 272
FT REPEAT 273
FT REPEAT 274
FT REPEAT 275
FT REPEAT 276
FT REPEAT 277
FT REPEAT 278
FT REPEAT 279
FT REPEAT 280
FT REPEAT 281
FT REPEAT 282
FT REPEAT 283
FT REPEAT 284
FT REPEAT 285
FT REPEAT 286
FT REPEAT 287
FT REPEAT 288
FT REPEAT 289
FT REPEAT 290
FT REPEAT 291
FT REPEAT 292
FT REPEAT 293
FT REPEAT 294
FT REPEAT 295
FT REPEAT 296
FT REPEAT 297
FT REPEAT 298
FT REPEAT 299
FT REPEAT 300
FT REPEAT 301
FT REPEAT 302
FT REPEAT 303
FT REPEAT 304
FT REPEAT 305
FT REPEAT 306
FT REPEAT 307
FT REPEAT 308
FT REPEAT 309
FT REPEAT 310
FT REPEAT 311
FT REPEAT 312
FT REPEAT 313
FT REPEAT 314
FT REPEAT 315
FT REPEAT 316
FT REPEAT 317
FT REPEAT 318
FT REPEAT 319
FT REPEAT 320
FT REPEAT 321
FT REPEAT 322
FT REPEAT 323
FT REPEAT 324
FT REPEAT 325
FT REPEAT 326
FT REPEAT 327
FT REPEAT 328
FT REPEAT 329
FT REPEAT 330
FT REPEAT 331
FT REPEAT 332
FT REPEAT 333
FT REPEAT 334
FT REPEAT 335
FT REPEAT 336
FT REPEAT 337
FT REPEAT 338
FT REPEAT 339
FT REPEAT 340
FT REPEAT 341
FT REPEAT 342
FT REPEAT 343
FT REPEAT 344
FT REPEAT 345
FT REPEAT 346
FT REPEAT 347
FT REPEAT 348
FT REPEAT 349
FT REPEAT 350
FT REPEAT 351
FT REPEAT 352
FT REPEAT 353
FT REPEAT 354
FT REPEAT 355
FT REPEAT 356
FT REPEAT 357
FT REPEAT 358
FT REPEAT 359
FT REPEAT 360
FT REPEAT 361
FT REPEAT 362
FT REPEAT 363
FT REPEAT 364
FT REPEAT 365
FT REPEAT 366
FT REPEAT 367
FT REPEAT 368
FT REPEAT 369
FT REPEAT 370
FT REPEAT 371
FT REPEAT 372
FT REPEAT 373
FT REPEAT 374
FT REPEAT 375
FT REPEAT 376
FT REPEAT 377
FT REPEAT 378
FT REPEAT 379
FT REPEAT 380
FT REPEAT 381
FT REPEAT 382
FT REPEAT 383
FT REPEAT 384
FT REPEAT 385
FT REPEAT 386
FT REPEAT 387
FT REPEAT 388
FT REPEAT 389
FT REPEAT 390
FT REPEAT 391
FT REPEAT 392
FT REPEAT 393
FT REPEAT 394
FT REPEAT 395
FT REPEAT 396
FT REPEAT 397
FT REPEAT 398
FT REPEAT 399
FT REPEAT 400
FT REPEAT 401
FT REPEAT 402
FT REPEAT 403
FT REPEAT 404
FT REPEAT 405
FT REPEAT 406
FT REPEAT 407
FT REPEAT 408
FT REPEAT 409
FT REPEAT 410
FT REPEAT 411
FT REPEAT 412
FT REPEAT 413
FT REPEAT 414
FT REPEAT 415
FT REPEAT 416
FT REPEAT 417
FT REPEAT 418
FT REPEAT 419
FT REPEAT 420
FT REPEAT 421
FT REPEAT 422
FT REPEAT 423
FT REPEAT 424
FT REPEAT 425
FT REPEAT 426
FT REPEAT 427
FT REPEAT 428
FT REPEAT 429
FT REPEAT 430
FT REPEAT 431
FT REPEAT 432
FT REPEAT 433
FT REPEAT 434
FT REPEAT 435
FT REPEAT 436
FT REPEAT 437
FT REPEAT 438
FT REPEAT 439
FT REPEAT 440
FT REPEAT 441
FT REPEAT 442
FT REPEAT 443
FT REPEAT 444
FT REPEAT 445
FT REPEAT 446
FT REPEAT 447
FT REPEAT 448
FT REPEAT 449
FT REPEAT 450
FT REPEAT 451
FT REPEAT 452
FT REPEAT 453
FT REPEAT 454
FT REPEAT 455
FT REPEAT 456
FT REPEAT 457
FT REPEAT 458
FT REPEAT 459
FT REPEAT 460
FT REPEAT 461
FT REPEAT 462
FT REPEAT 463
FT REPEAT 464
FT REPEAT 465
FT REPEAT 466
FT REPEAT 467
FT REPEAT 468
FT REPEAT 469
FT REPEAT 470
FT REPEAT 471
FT REPEAT 472
FT REPEAT 473
FT REPEAT 474
FT REPEAT 475
FT REPEAT 476
FT REPEAT 477
FT REPEAT 478
FT REPEAT 479
FT REPEAT 480
FT REPEAT 481
FT REPEAT 482
FT REPEAT 483
FT REPEAT 484
FT REPEAT 485
FT REPEAT 486
FT REPEAT 487
FT REPEAT 488
FT REPEAT 489
FT REPEAT 490
FT REPEAT 491
FT REPEAT 492
FT REPEAT 493
FT REPEAT 494
FT REPEAT 495
FT REPEAT 496
FT REPEAT 497
FT REPEAT 498
FT REPEAT 499
FT REPEAT 500
FT REPEAT 501
FT REPEAT 502
FT REPEAT 503
FT REPEAT 504
FT REPEAT 505
FT REPEAT 506
FT REPEAT 507
FT REPEAT 508
FT REPEAT 509
FT REPEAT 510
FT REPEAT 511
FT REPEAT 512
FT REPEAT 513
FT REPEAT 514
FT REPEAT 515
FT REPEAT 516
FT REPEAT 517
FT REPEAT 518
FT REPEAT 519
FT REPEAT 520
FT REPEAT 521
FT REPEAT 522
FT REPEAT 523
FT REPEAT 524
FT REPEAT 525
FT REPEAT 526
FT REPEAT 527
FT REPEAT 528
FT REPEAT 529
FT REPEAT 530
FT REPEAT 531
FT REPEAT 532
FT REPEAT 533
FT REPEAT 534
FT REPEAT 535
FT REPEAT 536
FT REPEAT 537
FT REPEAT 538
FT REPEAT 539
FT REPEAT 540
FT REPEAT 541
FT REPEAT 542
FT REPEAT 543
FT REPEAT 544
FT REPEAT 545
FT REPEAT 546
FT REPEAT 547
FT REPEAT 548
FT REPEAT 549
FT REPEAT 550
FT REPEAT 551
FT REPEAT 552
FT REPEAT 553
FT REPEAT 554
FT REPEAT 555
FT REPEAT 556
FT REPEAT 557
FT REPEAT 558
FT REPEAT 559
FT REPEAT 560
FT REPEAT 561
FT REPEAT 562
FT REPEAT 563
FT REPEAT 564
FT REPEAT 565
FT REPEAT 566
FT REPEAT 567
FT REPEAT 568
FT REPEAT 569
FT REPEAT 570
FT REPEAT 571
FT REPEAT 572
FT REPEAT 573
FT REPEAT 574
FT REPEAT 575
FT REPEAT 576
FT REPEAT 577
FT REPEAT 578
FT REPEAT 579
FT REPEAT 580
FT REPEAT 581
FT REPEAT 582
FT REPEAT 583
FT REPEAT 584
FT REPEAT 585
FT REPEAT 586
FT REPEAT 587
FT REPEAT 588
FT REPEAT 589
FT REPEAT 590
FT REPEAT 591
FT REPEAT 592
FT REPEAT 593
FT REPEAT 594
FT REPEAT 595
FT REPEAT 596
FT REPEAT 597
FT REPEAT 598
FT REPEAT 599
FT REPEAT 600
FT REPEAT 601
FT REPEAT 602
FT REPEAT 603
FT REPEAT 604
FT REPEAT 605
FT REPEAT 606
FT REPEAT 607
FT REPEAT 608
FT REPEAT 609
FT REPEAT 610
FT REPEAT 611
FT REPEAT 612
FT REPEAT 613
FT REPEAT 614
FT REPEAT 615
FT REPEAT 616
FT REPEAT 617
FT REPEAT 618
FT REPEAT 619
FT REPEAT 620
FT REPEAT 621
FT REPEAT 622
FT REPEAT 623
FT REPEAT 624
FT REPEAT 625
FT REPEAT 626
FT REPEAT 627
FT REPEAT 628
FT REPEAT 629
FT REPEAT 630
FT REPEAT 631
FT REPEAT 632
FT REPEAT 633
FT REPEAT 634
FT REPEAT 635
FT REPEAT 636
FT REPEAT 637
FT REPEAT 638
FT REPEAT 639
FT REPEAT 640
FT REPEAT 641
FT REPEAT 642
FT REPEAT 643
FT REPEAT 644
FT REPEAT 645
FT REPEAT 646
FT REPEAT 647
FT REPEAT 648
FT REPEAT 649
FT REPEAT 650
FT REPEAT 651
FT REPEAT 652
FT REPEAT 653
FT REPEAT 654
FT REPEAT 655
FT REPEAT 656
FT REPEAT 657
FT REPEAT 658
FT REPEAT 659
FT REPEAT 660
FT REPEAT 661
FT REPEAT 662
FT REPEAT 663
FT REPEAT 664
FT REPEAT 665
FT REPEAT 666
FT REPEAT 667
FT REPEAT 668
FT REPEAT 669
FT REPEAT 670
FT REPEAT 671
FT REPEAT 672
FT REPEAT 673
FT REPEAT 674
FT REPEAT 675
FT REPEAT 676
FT REPEAT 677
FT REPEAT 678
FT REPEAT 679
FT REPEAT 680
FT REPEAT 681
FT REPEAT 682
FT REPEAT 683
FT REPEAT 684
FT REPEAT 685
FT REPEAT 686
FT REPEAT 687
FT REPEAT 688
FT REPEAT 689
FT REPEAT 690
FT REPEAT 691
FT REPEAT 692
FT REPEAT 693
FT REPEAT 694
FT REPEAT 695
FT REPEAT 696
FT REPEAT 697
FT REPEAT 698
FT REPEAT 699
FT REPEAT 700
FT REPEAT 701
FT REPEAT 702
FT REPEAT 703
FT REPEAT 704
FT REPEAT 705
FT REPEAT 706
FT REPEAT 707
FT REPEAT 708
FT REPEAT 709
FT REPEAT 710
FT REPEAT 711
FT REPEAT 712
FT REPEAT 713
FT REPEAT 714
FT REPEAT 715
FT REPEAT 716
FT REPEAT 717
FT REPEAT 718
FT REPEAT 719
FT REPEAT 720
FT REPEAT 721
FT REPEAT 722
FT REPEAT 723
FT REPEAT 724
FT REPEAT 725
FT REPEAT 726
FT REPEAT 727
FT REPEAT 728
FT REPEAT 729
FT REPEAT 730
FT REPEAT 731
FT REPEAT 732
FT REPEAT 733
FT REPEAT 734
FT REPEAT 735
FT REPEAT 736
FT REPEAT 737
FT REPEAT 738
FT REPEAT 739
FT REPEAT 740
FT REPEAT 741
FT REPEAT 742
FT REPEAT 743
FT REPEAT 744
FT REPEAT 745
FT REPEAT 746
FT REPEAT 747
FT REPEAT 748
FT REPEAT 749
FT REPEAT 750
FT REPEAT 751
FT REPEAT 752
FT REPEAT 753
FT REPEAT 754
FT REPEAT 755
FT REPEAT 756
FT REPEAT 757
FT REPEAT 758
FT REPEAT 759
FT REPEAT 760
FT REPEAT 761
FT REPEAT 762
FT REPEAT 763
FT REPEAT 764
FT REPEAT 765
FT REPEAT 766
FT REPEAT 767
FT REPEAT 768
FT REPEAT 769
FT REPEAT 770
FT REPEAT 771
FT REPEAT 772
FT REPEAT 773
FT REPEAT 774
FT REPEAT 775
FT REPEAT 776
FT REPEAT 777
FT REPEAT 778
FT REPEAT 779
FT REPEAT 780
FT REPEAT 781
FT REPEAT 782
FT REPEAT 783
FT REPEAT 784
FT REPEAT 785
FT REPEAT 786
FT REPEAT 787
FT REPEAT 788
FT REPEAT 789
FT REPEAT 790
FT REPEAT 791
FT REPEAT 792
FT REPEAT 793
FT REPEAT 794
FT REPEAT 795
FT REPEAT 796
FT REPEAT 797
FT REPEAT 798
FT REPEAT 799
FT REPEAT 800
FT REPEAT 801
FT REPEAT 802
FT REPEAT 803
FT REPEAT 804
FT REPEAT 805
FT REPEAT 806
FT REPEAT 807
FT REPEAT 808
FT REPEAT 809
FT REPEAT 810
FT REPEAT 811
FT REPEAT 812
FT REPEAT 813
FT REPEAT 814
FT REPEAT 815
FT REPEAT 816
FT REPEAT 817
FT REPEAT 818
FT REPEAT 819
FT REPEAT 820
FT REPEAT 821
FT REPEAT 822
FT REPEAT 823
FT REPEAT 824
FT REPEAT 825
FT REPEAT 826
FT REPEAT 827
FT REPEAT 828
FT REPEAT 829
FT REPEAT 830
FT REPEAT 831
FT REPEAT 832
FT REPEAT 833
FT REPEAT 834
FT REPEAT 835
FT REPEAT 836
FT REPEAT 837
FT REPEAT 838
FT REPEAT 839
FT REPEAT 840
FT REPEAT 841
FT REPEAT 842
FT REPEAT 843
FT REPEAT 844
FT REPEAT 845
FT REPEAT 846
FT REPEAT 847
FT REPEAT 848
FT REPEAT 849
FT REPEAT 850
FT REPEAT 851
FT REPEAT 852
FT REPEAT 853
FT REPEAT 854
FT REPEAT 855
FT REPEAT 856
FT REPEAT 857
FT REPEAT 858
FT REPEAT 859
FT REPEAT 860
FT REPEAT 861
FT REPEAT 862
FT REPEAT 863
FT REPEAT 864
FT REPEAT 865
FT REPEAT 866
FT REPEAT 867
FT REPEAT 868
FT REPEAT 869
FT REPEAT 870
FT REPEAT 871
FT REPEAT 872
FT REPEAT 873
FT REPEAT 874
FT REPEAT 875
FT REPEAT 876
FT REPEAT 877
FT REPEAT 878
FT REPEAT 879
FT REPEAT 880
FT REPEAT 881
FT REPEAT 882
FT REPEAT 883
FT REPEAT 884
FT REPEAT 885
FT REPEAT 886
FT REPEAT 887
FT REPEAT 888
FT REPEAT 889
FT REPEAT 890
FT REPEAT 891
FT REPEAT 892
FT REPEAT 893
FT REPEAT 894
FT REPEAT 895
FT REPEAT 896
FT REPEAT 897
FT REPEAT 898
FT REPEAT 899
FT REPEAT 900
FT REPEAT 901
FT REPEAT 902
FT REPEAT 903
FT REPEAT 904
FT REPEAT 905
FT REPEAT 906
FT REPEAT 907
FT REPEAT 908
FT REPEAT 909
FT REPEAT 910
FT REPEAT 911
FT REPEAT 912
FT REPEAT 913
FT REPEAT 914
FT REPEAT 915
FT REPEAT 916
FT REPEAT 917
FT REPEAT 918
FT REPEAT 919
FT REPEAT 920
FT REPEAT 921
FT REPEAT 922
FT REPEAT 923
FT REPEAT 924
FT REPEAT 925
FT REPEAT 926
FT REPEAT 927
FT REPEAT 928
FT REPEAT 929
FT REPEAT 930
FT REPEAT 931
FT REPEAT 932
FT REPEAT 933
FT REPEAT 934
FT REPEAT 935
FT REPEAT 936
FT REPEAT 937
FT REPEAT 938
FT REPEAT 939
FT REPEAT 940
FT REPEAT 941
FT REPEAT 942
FT REPEAT 943
FT REPEAT 944
FT REPEAT 945
FT REPEAT 946
FT REPEAT 947
FT REPEAT 948
FT REPEAT 949
FT REPEAT 950
FT REPEAT 951
FT REPEAT 952
FT REPEAT 953
FT REPEAT 954
FT REPEAT 955
FT REPEAT 956
FT REPEAT 957
FT REPEAT 958
FT REPEAT 959
FT REPEAT 960
FT REPEAT 961
FT REPEAT 962
FT REPEAT 963
FT REPEAT 964
FT REPEAT 965
FT REPEAT 966
FT REPEAT 967
FT REPEAT 968
FT REPEAT 969
FT REPEAT 970
FT REPEAT 971
FT REPEAT 972
FT REPEAT 973
FT REPEAT 974
FT REPEAT 975
FT REPEAT 976
FT REPEAT 977
FT REPEAT 978
FT REPEAT 979
FT REPEAT 980
FT REPEAT 981
FT REPEAT 982
FT REPEAT 983
FT REPEAT 984
FT REPEAT 985
FT REPEAT 986
FT REPEAT 987
FT REPEAT 988
FT REPEAT 989
FT REPEAT 990
FT REPEAT 991
FT REPEAT 992
FT REPEAT 993
FT REPEAT 994
FT REPEAT 995
FT REPEAT 996
FT REPEAT 997
FT REPEAT 998
FT REPEAT 999
FT REPEAT 1000
FT REPEAT 1001
FT REPEAT 1002
FT REPEAT 1003
FT REPEAT 1004
FT REPEAT 1005
FT REPEAT 1006
FT REPEAT 1007
FT REPEAT 1008
FT REPEAT 1009
FT REPEAT 1010
FT REPEAT 1011
FT REPEAT 1012
FT REPEAT 1013
FT REPEAT 1014
FT REPEAT 1015
FT REPEAT 1016
FT REPEAT 1017
FT REPEAT 1018
FT REPEAT 1019
FT REPEAT 1020
FT REPEAT 1021
FT REPEAT 1022
FT REPEAT 1023
FT REPEAT 1024
FT REPEAT 1025
FT REPEAT 1026
FT REPEAT 1027
FT REPEAT 1028
FT REPEAT 1029
FT REPEAT 1030
FT REPEAT 1031
FT REPEAT 1032
FT REPEAT 1033
FT REPEAT 1034
FT REPEAT 1035
FT REPEAT 1036
FT REPEAT 1037
FT REPEAT 1038
FT REPEAT 1039
FT REPEAT 1040
FT REPEAT 1041
FT REPEAT 1042
FT REPEAT 1043
FT REPEAT 1044
FT REPEAT 1045
FT REPEAT 1046
FT REPEAT 1047
FT REPEAT 1048
FT REPEAT 1049
FT REPEAT 1050
FT REPEAT 1051
FT REPEAT 1052
FT REPEAT 1053
FT REPEAT 1054
FT REPEAT 1055
FT REPEAT 1056
FT REPEAT 1057
FT REPEAT 1058
FT REPEAT 1059
FT REPEAT 1060
FT REPEAT 1061
FT REPEAT 1062
FT REPEAT 1063
FT REPEAT 1064
FT REPEAT 1065
FT REPEAT 1066
FT REPEAT 1067
FT REPEAT 1068
FT REPEAT 1069
FT REPEAT 1070
FT REPEAT 1071
FT REPEAT 1072
FT REPEAT 1073
FT REPEAT 1074
FT REPEAT 1075
FT REPEAT 1076
FT REPEAT 1077
FT REPEAT 1078
FT REPEAT 1079
FT REPEAT 1080
FT REPEAT 1081
FT REPEAT 1082
FT REPEAT 1083
FT REPEAT 1084
FT REPEAT 1085
FT REPEAT 1086
FT REPEAT 1087
FT REPEAT 1088
FT REPEAT 1089
FT REPEAT 1090
FT REPEAT 1091
FT REPEAT 1092
FT REPEAT 1093
FT REPEAT 1094
FT REPEAT 1095
FT REPEAT 1096
FT REPEAT 1097
FT REPEAT 1098
FT REPEAT 1099
FT REPEAT 1100
FT REPEAT 1101
FT REPEAT 1102
FT REPEAT 1103
FT REPEAT 1104
FT REPEAT 1105
FT REPEAT 1106
FT REPEAT 1107
FT REPEAT 1108
FT REPEAT 1109
FT REPEAT 1110
FT REPEAT 1111
FT REPEAT 1112
FT REPEAT 1113
FT REPEAT 1114
FT REPEAT 1115
FT REPEAT 1116
FT REPEAT 1117
FT REPEAT 1118
FT REPEAT 1119
FT REPEAT 1120
FT REPEAT 1121
FT REPEAT 1122
FT REPEAT 1123
FT REPEAT 1124
FT REPEAT 1125
FT REPEAT 1126
FT REPEAT 1127
FT REPEAT 1128
FT REPEAT 1129
FT REPEAT 1130
FT REPEAT 1131
FT REPEAT 1132
FT REPEAT 1133
FT REPEAT 1134
FT REPEAT 1135
FT REPEAT 1136
FT REPEAT 1137
FT REPEAT 1138
FT REPEAT 1139
FT REPEAT 1140
FT REPEAT 1141
FT REPEAT 1142
FT REPEAT 1143
FT REPEAT 1144
FT REPEAT 1145
FT REPEAT 1146
FT REPEAT 1147
FT REPEAT 1148
FT REPEAT 1149
FT REPEAT 1150
FT REPEAT 1151
FT REPEAT 1152
FT REPEAT 1153
FT REPEAT

FT SITE 1659 1659 CALCIUM ION SELECTIVITY AND PERMEABILITY
 FT MOD RES 1725 1725 (BY SIMILARITY).
 FT CA BIND 1743 1754 PHOSPHORYLATION (BY PKA) (POTENTIAL).
 FT CARBOHYD 237 1237 BY SIMILARITY.
 FT CARBOHYD 1559 1359 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 601 601 P -> L (IN 75).
 SQ SEQUENCE 2164 AA; 246978 MW; E9A80AC5B4758B8C CRC64;
 Query Match 72.7%; Score 32; DB 1; Length 2164;
 Best Local Similarity 55.6%; Pred. No. 1.1e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ISEMFLOIT 9
 Db 485 MSEMFLKOT 493
 RESULT 6
 CCAA_RAT STANDARD; PRT: 2212 AA.
 AC P54282; 001541;
 DT 01-OCT-1996 (Rel. 34; Created)
 DT 01-OCT-1996 (Rel. 34; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)
 DE Voltage-dependent P/Q-type calcium channel alpha-1a subunit (calcium
 channel, L type, alpha-1 polypeptide, isoform 4) (Brain calcium
 channel 1) (B1) (RAT brain class A) (RBA-1).
 GN CACNA1A OR CACNA1A4 OR CACNA4 OR CACNA3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE-Brain;
 RX MEDLINE=9128516; PubMed=1648226;
 RA Starr T.V.B., Fryetay W., Shutch T.P.;
 RT "Primary structure of a calcium channel that is highly expressed in
 the rat cerebellum."
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5621-5625 (1991).
 RN [2]
 RP SEQUENCE OF 1435-1667 FROM N.A. (CLONE RBA-65).
 RX MEDLINE=90239020; PubMed=1692134;
 RA Shutch T.P., Leonard J.P., Gilbert M.M., Lester H.A., Davidson N.;
 RT "Rat brain expresses a heterogeneous family of calcium channels."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3391-3395 (1990).
 RN [3]
 RP SEQUENCE OF 1354-1659 FROM N.A. (ISOFORM RKC8).
 RC TISSUE-Kidney;
 RX MEDLINE=93066265; PubMed=1279681;
 RA Yu A.S.L., Hebert S.C., Brenner B.M., Lytton J.;
 RT "Molecular characterization and nephron distribution of a family of
 transcripts encoding the pore-forming subunit of Ca²⁺ channels in the
 kidney."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10494-10498 (1992).
 RL -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
 ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
 IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
 CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
 CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
 GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/Q-TYPE CALCIUM
 CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
 ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-
 IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO
 DIHYDROPIRIDINES (DHP), AND OMEGA-CONOTOXIN-GYIA (OMEGA-CTX-
 GYIA).
 -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
 COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
 IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
 FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
 SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
 CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA

CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC; PURKINE CELLS CONTAIN
 CC PREDOMINANTLY P-TYPE VSCC, THE Q-TYPE BEING A PROMINENT CALCIUM
 CC CURRENT IN CEREbellar GRANULE CELLS. ALSO FOUND IN HEART, IN
 CC KIDNEY DISTAL CONVOLUTED TUBULE (DCT), AND IN PITUITARY.
 CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
 CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
 CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
 CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
 CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
 CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M64373; AAA0806.1; -.
 CC EMBL: M99222; AAA0896.1; -.
 CC InterPro: IPR000637; AT hook.
 CC InterPro: IPR001682; Ca/Na pore.
 CC InterPro: IPR002077; Ca channel.
 CC InterPro: IPR002111; Cat channel.
 CC InterPro: IPR000636; M+channel_nlg.
 CC Pfam: PF00520; ion trans; 4.
 CC PRINTS: PR00167; CACHANNEL.
 CC SMART: SM00384; AT hook; 1.
 DR Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 KM Calcium channel; Glycoprotein; Repeat; Multigene family;
 KM Calcium-binding; Phosphorylation; Alternative splicing.
 KM REPEAT 87 365
 FT REPEAT 1.
 FT REPEAT 11.
 FT REPEAT 1182 1465
 FT REPEAT 1502 1765
 FT DOMAIN 1 100
 FT TRANSSEM 101 119
 FT DOMAIN 120 138
 FT TRANSSEM 139 156
 FT DOMAIN 157 168
 FT TRANSSEM 169 184
 FT DOMAIN 185 192
 FT TRANSSEM 193 211
 FT DOMAIN 212 230
 FT TRANSSEM 231 250
 FT DOMAIN 251 337
 FT TRANSSEM 338 362
 FT DOMAIN 363 489
 FT TRANSSEM 490 509
 FT DOMAIN 510 523
 FT TRANSSEM 524 543
 FT DOMAIN 544 551
 FT TRANSSEM 552 570
 FT DOMAIN 571 580
 FT TRANSSEM 581 599
 FT DOMAIN 600 618
 FT TRANSSEM 619 638
 FT DOMAIN 639 691
 FT TRANSSEM 692 716
 FT DOMAIN 717 1190
 FT TRANSSEM 1191 1214
 FT DOMAIN 1215 1231
 FT TRANSSEM 1232 1251
 FT DOMAIN 1252 1258
 FT TRANSSEM 1259 1293
 FT DOMAIN 1293 1293
 FT TRANSSEM 1294 1311
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M64373; AAA0806.1; -.
 CC EMBL: M99222; AAA0896.1; -.
 CC InterPro: IPR000637; AT hook.
 CC InterPro: IPR001682; Ca/Na pore.
 CC InterPro: IPR002077; Ca channel.
 CC InterPro: IPR002111; Cat channel.
 CC InterPro: IPR000636; M+channel_nlg.
 CC Pfam: PF00520; ion trans; 4.
 CC PRINTS: PR00167; CACHANNEL.
 CC SMART: SM00384; AT hook; 1.
 DR Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 KM Calcium channel; Glycoprotein; Repeat; Multigene family;
 KM Calcium-binding; Phosphorylation; Alternative splicing.
 KM REPEAT 87 365
 FT REPEAT 1.
 FT REPEAT 11.
 FT REPEAT 1182 1465
 FT REPEAT 1502 1765
 FT DOMAIN 1 100
 FT TRANSSEM 101 119
 FT DOMAIN 120 138
 FT TRANSSEM 139 156
 FT DOMAIN 157 168
 FT TRANSSEM 169 184
 FT DOMAIN 185 192
 FT TRANSSEM 193 211
 FT DOMAIN 212 230
 FT TRANSSEM 231 250
 FT DOMAIN 251 337
 FT TRANSSEM 338 362
 FT DOMAIN 363 489
 FT TRANSSEM 490 509
 FT DOMAIN 510 523
 FT TRANSSEM 524 543
 FT DOMAIN 544 551
 FT TRANSSEM 552 570
 FT DOMAIN 571 580
 FT TRANSSEM 581 599
 FT DOMAIN 600 618
 FT TRANSSEM 619 638
 FT DOMAIN 639 691
 FT TRANSSEM 692 716
 FT DOMAIN 717 1190
 FT TRANSSEM 1191 1214
 FT DOMAIN 1215 1231
 FT TRANSSEM 1232 1251
 FT DOMAIN 1252 1258
 FT TRANSSEM 1259 1293
 FT DOMAIN 1293 1293
 FT TRANSSEM 1294 1311

```

FT DOMAIN 1312 1330 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1331 1350 S5 OF REPEAT III (POTENTIAL).
FT DOMAIN 1351 1437 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1438 1462 S6 OF REPEAT III (POTENTIAL).
FT DOMAIN 1463 1518 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1519 1537 S1 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1538 1551 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1552 1573 S2 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1574 1580 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1581 1600 S3 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1601 1607 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1608 1626 S4 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1627 1645 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1646 1665 S5 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1666 1737 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1738 1763 S6 OF REPEAT IV (POTENTIAL).
FT TRANSMEM 1764 2212 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 13 20 POLY-GLY.
FT TRANSMEM 729 734 POLY-GLU.
FT TRANSMEM 1155 1158 POLY-HIS.
FT TRANSMEM 2162 2171 BINDING TO THE BETA SUBUNIT (BY
FT TRANSMEM 385 402 SIMILARITY).
FT TRANSMEM 320 320 CALCIUM ION SELECTIVITY AND PERMEABILITY
FT TRANSMEM 670 670 (BY SIMILARITY).
FT TRANSMEM 1411 1411 CALCIUM ION SELECTIVITY AND PERMEABILITY
FT TRANSMEM 1707 1707 (BY SIMILARITY).
FT TRANSMEM 1773 1773 CALCIUM ION SELECTIVITY AND PERMEABILITY
FT TRANSMEM 1791 1802 (BY SIMILARITY).
FT TRANSMEM 285 285 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT TRANSMEM 1607 1607 N-LINKED (GLYCAC. . .) (POTENTIAL).
FT TRANSMEM 1602 1602 N-LINKED (GLYCAC. . .) (POTENTIAL).
FT TRANSMEM 2212 2212 G -> GNP (IN ISOFORM RECB).
FT TRANSMEM 25125 25125 MW; DFB07A0F533AFA52 CRC64;
SO SPOUNCE

Query Match 72.7%; Score 32; DB 1; Length 2212;
Best local Similarity 55.6%; Pred. No. 1.1e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMPLOY 9
DB 533 WSEMPLOY 541

RESULT 7
CCAA RABIT STANDARD; PRT; 2424 AA.
AC P27884; P27883;
AT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Voltage-dependent P/Q-type calcium channel alpha-1a subunit (Calcium
channel, L type, alpha-1 polypeptide isoform 4) (Brain calcium channel
DE 1) (B1).
GN CACNA1A OR CACNA1A4 OR CACNA4 OR CACNA.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC MEDLINE=91187110; PubMed=1849233;
RA Mori Y., Fiedrich T., Kim M.-S., Miyami A., Nakai J., Ruth P.,
RA Bosse E., Hofman F., Flockezy V., Furutachi T., Motoshida K.,
RA Imoto K., Tanabe T., Numa S.;
RT "Primary structure and functional expression from complementary DNA
RT of a brain calcium channel."
RL Nature 350:398-402 (1991).
[2]

```

```

RP BETA-SUBUNIT BINDING DOMAIN, AND MUTAGENESIS.
RX MEDLINE=94150724; PubMed=7509046;
RA Pragnell M., de Waard M., Mori Y., Tanabe T., Snutch T.P.,
RA Campbell K.P.;
RT "Calcium channel beta-subunit binds to a conserved motif in the I-II
RT cytoplasmic linker of the alpha 1-subunit."
RL Nature 368:67-70 (1994).
CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
CC GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/Q-TYPE CALCIUM
CC CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
CC ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-
CC IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO
CC DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-
CC GVIA).
CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC here), CBP101/CBP109, CBP103 and CBP107, are produced by
CC alternative splicing.
CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC. PURKINJE CELLS CONTAIN
CC PREDOMINANTLY P-TYPE VSCC. THE Q-TYPE BEING A PROMINENT CALCIUM
CC CURRENT IN CEREBELLAR GRANULE CELLS.
CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X57477; CAA40715.1; -.
DR EMBL, X57689; CAA40872.1; -.
DR EMBL, X57476; CAA40714.1; -.
DR EMBL, X57688; CAA40871.1; -.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR001682; Ca/Na pore.
DR InterPro; IPR002077; Ca_channel.
DR InterPro; IPR002111; Cat_channel_Typl.
DR InterPro; IPR000636; M+channel_nlg.
DR Pfam; PF00520; Ion trans; 4.
DR PRINTS; PR00167; CACHANNEL.
DR SMART; SM00384; AT hook; 1.
DR Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
DR Calcium channel; Glycoprotein; Repeat; Multigene family;
KW Calcium-binding; Phosphorylation; Alternative splicing.
FT REPEAT 473 717 I.
FT REPEAT 1240 1523 III.
FT REPEAT 1560 1823 IV.
FT TRANSMEM 99 98 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 117 117 S1 OF REPEAT I (POTENTIAL).
FT TRANSMEM 118 135 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 136 155 S2 OF REPEAT I (POTENTIAL).
FT TRANSMEM 156 167 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 168 185 S3 OF REPEAT I (POTENTIAL).

```

FT	TRANSM	186	190	EXTRACELLULAR (POTENTIAL).
FT	TRANSM	191	209	S4 OF REPEAT I (POTENTIAL).
FT	DOMAIN	210	228	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	229	248	S5 OF REPEAT I (POTENTIAL).
FT	DOMAIN	249	335	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	336	360	S6 OF REPEAT I (POTENTIAL).
FT	DOMAIN	361	487	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	488	506	S1 OF REPEAT II (POTENTIAL).
FT	DOMAIN	507	521	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	522	549	S2 OF REPEAT II (POTENTIAL).
FT	DOMAIN	542	541	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	550	568	S3 OF REPEAT II (POTENTIAL).
FT	DOMAIN	569	578	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	579	597	S4 OF REPEAT II (POTENTIAL).
FT	DOMAIN	598	616	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	617	636	S5 OF REPEAT II (POTENTIAL).
FT	DOMAIN	637	689	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	690	714	S6 OF REPEAT II (POTENTIAL).
FT	DOMAIN	715	1253	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1254	1272	S1 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1273	1288	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1289	1308	S2 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1309	1320	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1321	1339	S3 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1340	1350	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1351	1369	S4 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1370	1388	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1389	1408	S5 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1409	1495	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1496	1520	S6 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1521	1575	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1576	1604	S1 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1605	1609	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1610	1629	S2 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1630	1637	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1638	1656	S3 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1657	1665	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1666	1684	S4 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1685	1703	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1704	1723	S5 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1724	1795	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1796	1820	S6 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1821	2424	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	13	18	POLY-GLY.
FT	DOMAIN	727	732	BINDING TO THE BETA SUBUNIT.
FT	DOMAIN	1004	1010	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	DOMAIN	1012	1017	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	DOMAIN	2219	2227	POLY-ARG.
FT	DOMAIN	2242	2246	POLY-HIS.
FT	DOMAIN	2268	2287	POLY-ARG.
FT	DOMAIN	2298	2301	POLY-GLY.
FT	DOMAIN	2372	2377	POLY-PRO.
FT	DOMAIN	2411	2416	POLY-GLY.
FT	DOMAIN	383	400	BINDING TO THE BETA SUBUNIT.
FT	DOMAIN	318	318	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	SITE	668	668	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	SITE	1469	1469	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	SITE	1765	1765	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	MOD_RES	1831	1831	PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT	CA_BIND	1849	1860	BY SIMILARITY.
FT	CARBOHYD	283	283	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1665	1665	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC	772	1051	MISSING (IN ISOPOM CBP107).
FT	VARSPPLIC	772	1120	MISSING (IN ISOPOM CBP103).
FT	VARSPPLIC	1857	1884	LYSDNYMGRHPGICGKCHHRAC -> HYDNYMSLLRHPGICGKCHHRAC (IN ISOPOM CBP101).
FT	VARSPPLIC	2230	2273	RGGRGYSRVSARRRRRGVAVRPAAPALAHARRARAFARL -> PVAADKRRGVPQDRDGHRRARQRQWSSSPS

FT	VARSPPLIC	2274	2424		EGREHTHRO (IN ISOFORM BI-1).
FT	VARIANT	419	419		MISSING (IN ISOFORM BI-1).
FT	VARIANT	877	877		MISSING (IN ISOFORM CBP15).
FT	VARIANT	1104	1104		A -> T (IN ISOFORM CBS).
FT	MUTAGEN	385	386		S -> N (IN ISOFORM CBS).
FT	MUTAGEN	389	389		E-S: REDUCED BETA-SUBUNIT INTERACTION.
FT	MUTAGEN	392	392		L->S: REDUCED BETA-SUBUNIT INTERACTION.
FT	MUTAGEN	400	400		T->S: REDUCED BETA-SUBUNIT INTERACTION.
FT	SEQUENCE	2424 AA;	273228 MW;	F7CCAD0AB4B5604 CRC64;	INTERACTION.
SO	Query Match	72.7%;	Score 32;	DB 1;	Length 2424;
	Best Local Similarity	55.6%;	Pred. No. 1.2e+02;		
	Matches	5;	Conservative	4;	Mismatches 0; Indels 0; Gaps 0;
Ox	1 ISEMPLQTY 9				
Db	531 MSBMPKMY 539				
	RESULT 8				
	CCAA HUMAN				
ID	CCAA_HUMAN	STANDARD;	PRT;	2505 AA.	
AC	000555:	Q92690; Q16290; Q99790; Q99791; Q99792; Q99793; P78510;			
AC	P78510				
DT	15-JUL-1999 (Rel. 38, Created)				
DT	15-JUN-2002 (Rel. 38, Last sequence update)				
DE	Voltage-dependent P/Q-type calcium channel alpha-1A subunit (Calcium channel, L type, alpha-1 polypeptide isoform 4) (Brain calcium channel I) (BI).				
GN	CACNA1A OR CACNL1A4 OR CAC4A OR CACN3.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCHI_TaxId=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORMS BI-1-GGCAG/1A-1 AND BI-1/1A-2).				
RC	TISSUE=Neuron;				
RX	MEDLINE=99158614; PubMed=10049321;				
RA	Hans M., Urrutia A., Deal C., Brust P.F., Stauderman K., Ellis S.B., Harbold M.M., Johnson E.C., Williams M.E.;				
RT	"structural elements in domain IV that influence biophysical and pharmacological properties of human alpha1A-containing high-voltage-activated calcium channels.";				
RL	Biophys. J. 76:1384-1400(1999).				
RP	SEQUENCE FROM N.A. (ISOFORM BI-1(V1)), AND VARIANTS FRM.				
RC	TISSUE=Cerebellum;				
RX	MEDLINE=97053792; PubMed=8898206;				
RA	Ophoff R.A., Terwindt G.W., Vergouwe M.N., van Bijl R., Oefner P.J., Hofman S.M.G., Lamerdin J.E., Moirewetsier H.W., Bulman D.B., Ferrari M., Haan J., Lindhout D., van Ommen G.-D.B., Hofker M.H., Ferrari M.D., Frances R.R.;				
RT	"familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca2+ channel gene CACNL1A4.";				
RL	Cell 87:543-552(1996).				
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.				
RC	TISSUE=Brain;				
RX	MEDLINE=97141920; PubMed=8988170;				
RA	Zuchenko O., Bailey J., Bonnen P.E., Ashizawa T., Stockton D.W., Anos C., Dobyns W.B., Subramony S.H., Zoghbi H.Y., Lee C.C.;				
RT	"autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel.";				
RL	Nat. Genet. 15:62-69(1997).				
RP	SEQUENCE OF 1233-1651 FROM N.A.				
RA	Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W., Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S., Phan H., Velasco N., Do L., Regala W., Terry A., Gaines D.,				

RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RA Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1693-1807 FROM N.A.
RC TISSUE=Lung carcinoma;
RX MEDLINE=95123419; PubMed=782133;
RA Barry E.L.R., Vighione M.P., Kim Y.I., Froehner S.C.;
RT "Expression and antibody inhibition of P-type calcium channels in
RT human small-cell lung carcinoma cells.";
RL J. Neurosci. 15:274-283(1995).
RN [6]
RP SEQUENCE OF 2038-2258 FROM N.A.
RC TISSUE=Frontal cortex;
RX MEDLINE=96102310; PubMed=8525433;
RA Margolis R.L., Breschel T.S., Li S.H., Kidwai A.S., Antonarakis S.E.,
RA McInnis M.G., Ross C.A.;
RT "Characterization of cDNA clones containing CCA trinucleotide repeats
RT derived from human brain.";
RL Somat. Cell Mol. Genet. 21:279-284(1995).
CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
CC GIVES RISE TO P AND/OR O-TYPE CALCIUM CURRENTS. P/O-TYPE CALCIUM
CC CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
CC ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AACOTOXIN-
CC IYA (OMEGA-AGA-IYA). THEY ARE HOWEVER INSENSITIVE TO
CC DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-
CC GVIA).
CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS; B1-1/1A-2, B1-1(V1),
CC B1-1/GGAG/1A-1 (SHOWN HERE), B1-1(V1)-GGGAG, B1-1(V2), B1-1(V2)-
CC GGAG AND B1-1(V2,V3); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC; MAINLY FOUND IN CEREBELLUM,
CC CEREBRAL CORTEX, THALAMUS AND HYPOTHALAMUS. NO EXPRESSION IN
CC HEART, KIDNEY, LIVER OR MUSCLE. PURKINJE CELLS CONTAIN
CC PREDOMINANTLY P-TYPE VSCC. THE Q-TYPE BEING A PROMINENT CALCIUM
CC CURRENT IN CEREBELLAR GRANULE CELLS.
CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -1- POLYMORPHISM: THE POLY-GLN REGION OF CANALINA IS POLYMORPHIC; 6 TO
CC 17 REPEATS IN THE NORMAL POPULATION, EXPANDED TO ABOUT 21 TO 30
CC REPEATS IN SPINOCEREBELLAR ATAXIA 6 (SC6) PATIENTS. THERE SEEMS
CC TO BE A CORRELATION BETWEEN THE REPEAT NUMBER AND EARLIER ONSET OF
CC THE DISORDER.
CC -1- DISEASE: EXPANSION OF A CAG REPEAT IN THE CODING REGION OF CANALINA
CC IS THE CAUSE OF ONE FORM OF SPINOCEREBELLAR ATAXIA 6 (SC6). AN
CC AUTOSOMAL DOMINANT DISORDER CHARACTERIZED BY SLOWLY PROGRESSIVE
CC CEREBELLAR ATAXIA OF THE LIMBS AND GAIT, DYSARTHRIA, NYSTAGMUS,
CC AND MILD VIBRATORY AND PROPRIOCEPTIVE SENSORY LOSS. THESE SYMPTOMS
CC ARE PROBABLY EXPLAINED BY SEVERE LOSS OF CEREBELLAR PURKINJE
CC CELLS.
CC -1- DISEASE: DEFECTS IN CANALINA ARE THE CAUSE OF FAMILIAL HEMIPLEGIC
CC MIGRAINE (FHM), ALSO KNOWN AS MIGRAINE FAMILIAL HEMIPLEGIC 1
CC (MHP1), EPISODIC ATAXIA TYPE 2 (EA-2), ACETAZOLAMIDE-RESPONSIVE
CC HEREDITARY PAROXYSMAL CEREBELLAR ATAXIA (APCA), AND POSSIBLY OTHER
CC COMMON TYPES OF MIGRAINES. FHM, A RARE AUTOSOMAL DOMINANT SUBTYPE
CC OF MIGRAINE WITH AURA, IS ASSOCIATED WITH ICTAL HEMIPARESIS AND,

CC IN SOME FAMILIES, PROGRESSIVE CEREBELLAR ATROPHY. EA-2 IS ANOTHER
CC AUTOSOMAL DOMINANT PAROXYSMAL CEREBELLAR DISEASE, CHARACTERIZED BY
CC ACETAZOLAMIDE-RESPONSIVE ATTACKS OF CEREBELLAR ATAXIA AND
CC MIGRAINE-LIKE SYMPTOMS, INTERICTAL NYSTAGMUS, AND CEREBELLAR
CC ATROPHY.
CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004884; AAB61613.1; -;
CC EMBL; AF004883; AAB61612.1; -;
CC EMBL; X99897; CA68172.1; -;
CC EMBL; Z80115; -; NOT_ANNOTATED_CDS.
CC EMBL; Z80115; -; NOT_ANNOTATED_CDS.
CC EMBL; U79666; AAB64179.1; -;
CC EMBL; U79663; AAB49674.1; ALT_INIT.
CC EMBL; U79664; AAB49675.1; ALT_INIT.
CC EMBL; U79665; AAB49676.1; ALT_INIT.
CC EMBL; U79667; AAB49677.1; ALT_INIT.
CC EMBL; U79668; AAB49678.1; ALT_INIT.
CC EMBL; AC005305; AAC26839.1; -;
CC EMBL; S76537; AAB33068.1; -;
CC EMBL; U06702; -; NOT_ANNOTATED_CDS.
CC Gene; HGNC:1388; CANALINA.
CC MIM; 601011; -;
CC MIM; 183086; -;
CC MIM; 141500; -;
CC MIM; 108500; -;
CC InterPro; IPR001682; Ca/Na_pore.
CC InterPro; IPR002077; Ca_channel.
CC InterPro; IPR002111; Cat_channel_TripL.
CC InterPro; IPR000636; M-channel_nlg.
CC Pfam; PF00520; Ion_trans_4.
CC PRINTS; PR00167; CACANALIN.
CC KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
CC KW Calcium channel; Glycoprotein; Repeat; Multigene family; Polymorphism;
CC KW Calcium-binding; Phosphorylation; Alternative splicing; Polymorphism;
CC FT Disease mutation; Triplet repeat expansion.
CC FT REPEAT 85 363
CC FT REPEAT 473 717
CC FT REPEAT 1231 1514
CC FT REPEAT 1551 1814
CC FT DOMAIN 1 98
CC FT TRANSMEM 99 117
CC FT DOMAIN 118 135
CC FT TRANSMEM 136 155
CC FT TRANSMEM 156 167
CC FT DOMAIN 168 185
CC FT TRANSMEM 186 190
CC FT DOMAIN 191 209
CC FT TRANSMEM 210 228
CC FT DOMAIN 229 248
CC FT TRANSMEM 249 335
CC FT DOMAIN 336 360
CC FT TRANSMEM 361 487
CC FT DOMAIN 488 506
CC FT TRANSMEM 507 521
CC FT DOMAIN 522 541
CC FT TRANSMEM 542 549
CC FT DOMAIN 549 568
CC FT TRANSMEM 569 578
CC FT DOMAIN 579 597
CC FT TRANSMEM 598 616
CC FT DOMAIN 617 636
CC FT TRANSMEM 637 689
CC FT DOMAIN 690 714
CC FT TRANSMEM 714
CC -----
CC I. CYTOPLASMIC (POTENTIAL).
CC S1 OF REPEAT I (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC S2 OF REPEAT I (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC S3 OF REPEAT I (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC S4 OF REPEAT I (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC S5 OF REPEAT I (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC S6 OF REPEAT I (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC S1 OF REPEAT II (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC S2 OF REPEAT II (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC S3 OF REPEAT II (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC S4 OF REPEAT II (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC S5 OF REPEAT II (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC S6 OF REPEAT II (POTENTIAL).

FT DOMAIN 715 1242 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1243 1261 S1 OF REPEAT II (POTENTIAL).
 FT DOMAIN 1262 1277 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1278 1297 S2 OF REPEAT III (POTENTIAL).

Query Match
 Best Local Similarity 72.7%; Score 32; DB 1; Length 2505;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLQIY 9
 DB 531 MSEMFLQIY 539

RESULT 9

SODM_PALVU STANDARD; PRT; 144 AA.
 AC P28765.
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Superoxide dismutase [Mn], mitochondrial (EC 1.15.1.1) (Fragment).
 OS *Palinurus vulgaris* (European spiny lobster).
 OC *Palinurus vulgaris* (Arthropoda: Mandibulata); Pancrustacea; Crustacea;
 OC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata;
 OC Palinura; Palinuroidea; Palinuridae; Palinurus.
 OC NCBI_TaxID=6733;

SEQUENCE FROM N.A.
 MEDLINE=92211732; PubMed=1556751;
 Smith M.W., Doolittle R.F.;
 "A comparison of evolutionary rates of the two major kinds of
 superoxide dismutase."
 J. Mol. Evol. 34:175-184 (1992).
 CC -1- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems.
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -1- COFACTOR: Manganese (By similarity).
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 CC FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; X64063; CA45419.1; -
 DR PIR; S23659; S23659.
 DR HSSP; P04179; IAP6.
 DR InterPro; IPR001189; SODismutase.
 DR Pfam; PF00081; sodfe.1.
 DR Pfam; PF02777; sodfe.C1.1.
 DR Prodom; PD000475; SODismutase; 1.
 DR PROSITE; PS00088; SOD_MN; PARTIAL.
 KW Oxidoreductase; Manganese; Mitochondrion.
 FT NON_TER 1
 FT METAL 10
 FT METAL 58
 FT METAL 143
 FT NON_TER 144

SEQUENCE 144 AA; 15505 MW; F180CE523653686 CRC64;
 Query Match
 Best Local Similarity 70.5%; Score 31; DB 1; Length 144;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ISEMFLQIY 9
 DB 531 MSEMFLQIY 539

DB 2 ISEMFLQIY 10

RESULT 10
 VGLM_HSVSA STANDARD; PRT; 366 AA.
 AC 001017;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Glycoprotein M.
 GN 39.
 OS Herpesvirus saimiri (strain 11).
 OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OC NCBI_TaxID=10383;
 RN [1]

SEQUENCE FROM N.A.
 MEDLINE=9233668; PubMed=1321287;
 RA Aldrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
 RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
 RA Honess R.W.;
 "Primary structure of the herpesvirus saimiri genome."
 J. Virol. 66:5047-5058 (1992).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN M.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; X64346; CA45662.1; -
 DR PIR; A36810; OOBENS.
 DR InterPro; IPR000785; Herpes_glycop.
 DR Pfam; PF01528; Herpes_glycop.1.
 DR PRINTS; PR00333; HSVINTEGRMP.
 KW Transmembrane; Glycoprotein.

FT TRANSMEM 18 34 POTENTIAL.
 FT TRANSMEM 87 103 POTENTIAL.
 FT TRANSMEM 152 168 POTENTIAL.
 FT TRANSMEM 214 230 POTENTIAL.
 FT TRANSMEM 236 252 POTENTIAL.
 FT TRANSMEM 274 290 POTENTIAL.
 FT TRANSMEM 305 321 POTENTIAL.
 FT DOMAIN 342 351 ARG/LYS-RICH (BASIC).
 FT CARBOHYD 57 57 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 366 AA; 42182 MW; 4EAE0C3639E76463 CRC64;

Query Match
 Best Local Similarity 70.5%; Score 31; DB 1; Length 366;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ISEMFLQIY 9
 DB 259 MTEMFLQIY 267

RESULT 11
 Y084_MYCTU STANDARD; PRT; 491 AA.
 AC 053209;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein RV2484C.
 GN RV2484C OR MT2557 OR KMY008.40C.
 OS Mycobacterium tuberculosis.
 OS Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

Qy 1 ISEMFLQIY 9
 DB 259 MTEMFLQIY 267


```

OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=96295987; PubMed=9634220;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gents S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.B., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
(2)
RN SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Dwyer R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson M.C., Mayhew L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,
RA Biswal W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
CC -1- SIMILARITY: BELONGS TO THE UPF0089 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL021246; CAA16061.1; -
DR EMBL; AE007092; AAK46861.1; -
DR TIGR; MT2557; -
DR Tuberculist; RV2484C; -
DR InterPro; IPR004255; UPF0089.
DR Pfam; PF03007; UPF0089; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 491 AA; 52309 MW; CZACCC458ADF82DD CRC64;

Query Match 70.5%; Score 31; DB 1; Length 491;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 EMBLOIDY 9
Db 155 EMBLOIDY 161

RESULT 12
YEJA_ECOLI STANDARD; FRT; 604 AA.
ID YEJA_ECOLI
AC P33913; P76447;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Yeja precursor.
GN YEJA OR B2177.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

```

```

RA Riley M., Collado-Vides J., Glaeser J.D., Rode C.K., Mayhew G.F.,
RA Gregor U., Davis N.M., Kirkpatrick H.A., Goeden M.A., Rose D.O.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
(2)
RN SEQUENCE OF 193-604 FROM N.A.
RP STRAIN=K12 / BHB2600;
RC STRAIN=K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBD databases.
RN [3]
RP SEQUENCE OF 193-604 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alpa H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Mocomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Sivasubramanian S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horikuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
-1- SIMILARITY: SOME, TO H. INFLUENZAE HBPA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000307; AAC75238.1; ALT_INIT.
DR EMBL; U00008; AAA16375.1; -
DR EMBL; D90849; BAA15985.1; ALT_INIT.
DR Ecogen; EGI2037; Yeja.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 2.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 604 HYPOTHETICAL PROTEIN YEJA.
FT CONFLICT 193 195 YMK -> IRP (IN REF. 2).
FT CONFLICT 415 415 W -> G (IN REF. 2).
SQ SEQUENCE 604 AA; 69669 MW; E3D45FAF94F16B47 CRC64;

Query Match 70.5%; Score 31; DB 1; Length 604;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 SEMELOIDY 9
Db 381 SEVLOIDY 388

RESULT 13
APCE_SYNY4 STANDARD; PRT; 896 AA.
ID APCE_SYNY4
AC 002907;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Phycobillibone 100.5 kDa core-membrane linker polypeptide (L-CM 100.5)
DE (Anchor polypeptide).
OS Synecocystis sp. (strain PCC 6714).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1147;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93222481; PubMed=8467079;
RA Dimagno L.M., Haselkorn R.;
RT "Isolation and characterization of the genes encoding allophycocyanin

```



```

RT subunits and two linker proteins from Synechocystis 6714."
RL Plant Mol. Biol. 21:835-846(1993).
CC -1- FUNCTION: THIS PROTEIN IS POSTULATED TO ACT BOTH AS TERMINAL
CC ENERGY ACCEPTOR (BY ITS PHYCOBILIN-LIKE DOMAIN) AND AS A LINKER
CC POLYPEPTIDE (BY ITS REPEATS AND ARMS) THAT STABILIZES THE
CC PHYCOBILISOME CORE ARCHITECTURE.
CC -1- SUBUNIT: PHYCOBILISOMES OF THIS ORGANISM ARE COMPOSED OF A TWO
CC CYLINDER CORE, FROM WHICH SIX RODS RADIATE. THE CORE IS MAINLY
CC COMPOSED OF ALLOPHYCOCYANIN ALPHA AND BETA CHAINS AND OF MINOR
CC COMPONENTS.
CC -1- SUBCELLULAR LOCATION: ANCHORS THE PHYCOBILISOME PERPENDICULARLY
CC TO THE STROMAL SURFACE OF THE THYLAKOID MEMBRANE.
CC -1- SIMILARITY: THE REPEATED DOMAINS ARE SIMILAR TO THE N-TERMINAL
CC REGIONS OF PHYCOCYANIN ROD LINKER POLYPEPTIDES.
CC -1- SIMILARITY: THE PHYCOBILIN-LIKE DOMAINS ARE SIMILAR TO PHYCOBILINS
CC FROM VARIOUS SPECIES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; 102309; AAA69685.1; -.
DR HSSP; P00318; 1B33.
DR InterPro; IPR001297; PBS_linker_poly.
DR InterPro; IPR001659; Phycobillisme.
DR Pfam; PF00427; PBS_linker_poly; 3.
DR Pfam; PF00502; Phycobillisme; 1.
DR ProDom; PD000340; Phycobillisme; 2.
DR Phycobillisme; Electron transport; Photosynthesis; Repeat.
FT DOMAIN 233 280 ARM 1 (SPACING SEQUENCE) (BY SIMILARITY).
FT REPEAT 281 405 1.
FT DOMAIN 406 541 ARM 2 (SPACING SEQUENCE) (BY SIMILARITY).
FT REPEAT 542 664 11.
FT DOMAIN 665 737 ARM 3 (SPACING SEQUENCE) (BY SIMILARITY).
FT REPEAT 738 863 111.
SQ SEQUENCE 896 AA; 100460 MW; BDID14C2923E32A3 CRC64;

Query Match 70.5%; Score 31; DB 1; Length 896;
Best Local Similarity 62.5%; Pred. No. 67;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMPLQTY 9
Db 585 SELFLKLY 592

RESULT 14
PR16_YEAST STANDARD; PRT; 1071 AA.
AC P15936;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pre-mRNA splicing factor RNA helicase PRP16.
DE PRP16 OR YKR086W OR YKR406.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90182661; PubMed=2138057;
RA Burgess S., Couto J.R., Guthrie C.;
RT "A putative ATP binding protein influences the fidelity of
RT branchpoint recognition in yeast splicing."
RL Cell 60:705-717(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94262327; PubMed=8203164;

```

```

RA Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
RA Remacha M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;
RT "The complete sequence of an 18,002 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the HBS1, MRP-L20 and PRP16 genes,
RT and six new open reading frames."
RL Yeast 10:231-245(1994).
RN [3]
RP FUNCTION.
RX MEDLINE=91125476; PubMed=1825134;
RA Scherer B., Guthrie C.;
RT "PRP16 is an RNA-dependent ATPase that interacts transiently with the
RT spliceosome."
RL Nature 349:494-499(1991).
CC -1- FUNCTION: INFLUENCES THE FIDELITY OF BRANCHPOINT RECOGNITION IN
CC YEAST SPLICING. THIS IS RNA-DEPENDENT ATPASE WHICH IS ESSENTIAL
CC FOR VIABILITY. IT MAY MEDIATE ONE OF THE MANY ATP-REQUIRING STEPS
CC OF SPLICOSOME ASSEMBLY AND THAT ACCURACY OF BRANCHPOINT
CC RECOGNITION MAY BE COUPLED TO ATP BINDING AND/OR HYDROLYSIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. BINDS TO THE SPLICOSOME.
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAD
CC SUBFAMILY. PRP16 ORTHOLOG.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M31524; AAA34911.1; -.
DR EMBL; 227116; CAA81637.1; -.
DR EMBL; 228311; CAA82165.1; -.
DR PIR; S38164; S38164.
DR PIR; S39127; S39127.
DR SGD; S0001794; PRP16.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAD_box.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00690; DEAD_ATP_HELICASE; 1.
DR mRNA processing; mRNA splicing; Helicase; ATP-binding;
KM Nuclear protein.
FT NP_BIND 373 380 ATP (BY SIMILARITY).
FT SITE 473 476 DEAD_BOX.
FT MUTAGEN 386 386 Y->D: SUPPRESSOR PHENOTYPE.
FT CONFLICT 698 698 A -> R (IN REF. 1).
SQ SEQUENCE 1071 AA; 121652 MW; EB7D4C4A18F5D4E1 CRC64;

Query Match 70.5%; Score 31; DB 1; Length 1071;
Best Local Similarity 55.6%; Pred. No. 81;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SEMPLQTY 9
Db 583 LOEXFLQY 591

RESULT 15
MSP1_PLAFK STANDARD; PRT; 1630 AA.
ID MSP1_PLAFK
AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merzoite surface protein 1 precursor (merozoite surface antigens)
DE (PMMSA) (P190).
OS Plasmodium falciparum (isolate KI / Thailand).
GN MSP-1.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;

```

```
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=66136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RT "Polymorphism of the precursor for the major surface antigens of
RL Plasmodium falciparum merozoites: studies at the genetic level.";
EMBO J. 4:3823-3829(1985).
RN [2]
RP REVISIONS. SEQUENCE FROM N.A.
RA Pan W., Tolle R., Bujard H.
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03371; CAA27070.1; -.
DR PIR; A25120; SAZOK1.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR Malaria; Merozoite; Polyporein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.
FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 759 759 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 955 955 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1630 AA; 187289 MW; ADHDEG3CE0A46322 CRC64;

Query Match 70.5%; Score 31; DB 1; Length 1630;
Best Local Similarity 44.4%; Pred. No. 1.3e+02;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
:|:|:|:|:|
Db 838 LSOLFMEIY 846
```

Search completed: May 1, 2003, 23:10:57
Job time : 3.32143 secs

Thu May 8 16:14:15 2003

us-09-658-621b-26.rapb

Page 1

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: May 1, 2003, 23:12:56 ; Search time 0.571429 Seconds

(Without alignments)
1359.015 Million cell updates/sec

Title: US-09-658-621b-26

Perfect score: 44

Sequence: 1 ISEMFLQIY 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328255 seqs, 8628685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCT05_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	321	10	US-09-925-301-861
2	44	100.0	1255	9	US-09-996-069-10
3	41	93.2	508	9	US-10-057-136-20
4	31	70.5	1639	9	US-10-087-464-10
5	30	68.2	207	9	US-09-963-959-7
6	30	68.2	272	9	US-10-060-019-27
7	30	68.2	815	9	US-10-060-019-24
8	29	65.9	44	9	US-09-972-970-5
9	29	65.9	69	9	US-10-091-504-1005
10	29	65.9	69	10	US-09-764-869-1005
11	29	65.9	485	10	US-09-925-300-1664
12	28	63.6	111	9	US-10-078-090-209
13	28	63.6	261	9	US-10-166-087-34
14	28	63.6	299	9	US-09-510-332-1
15	28	63.6	299	10	US-09-193-634-35
16	28	63.6	332	9	US-10-043-487-11
17	28	63.6	349	10	US-09-804-551B-4
18	28	63.6	353	10	US-09-841-132-510
19	28	63.6	410	9	US-09-987-021-3

20	28	63.6	410	10	US-09-957-485-3
21	28	63.6	662	12	US-10-036-328A-8
22	28	63.6	714	10	US-09-818-264-2
23	28	63.6	733	12	US-10-036-328A-4
24	28	63.6	829	9	US-10-102-806-602
25	28	63.6	1029	9	US-10-033-245-22
26	28	63.6	1029	9	US-10-033-223-22
27	28	63.6	1029	9	US-10-033-167-22
28	28	63.6	1029	9	US-10-033-244-22
29	28	63.6	1029	9	US-10-033-435-22
30	28	63.6	1029	9	US-10-032-990-22
31	28	63.6	1029	9	US-09-929-769-7
32	28	63.6	1029	9	US-10-032-996-22
33	28	63.6	1029	9	US-10-033-396-22
34	28	63.6	1029	12	US-10-033-246-22
35	28	63.6	1029	12	US-10-033-301-22
36	28	63.6	1029	12	US-10-033-326-22
37	28	63.6	1138	12	US-10-036-328A-6
38	28	63.6	1210	12	US-10-036-328A-2
39	27	61.4	61	9	US-10-092-154-566
40	27	61.4	61	10	US-09-764-847-566
41	27	61.4	67	10	US-09-729-674-166
42	27	61.4	84	9	US-09-764-891-5152
43	27	61.4	88	9	US-10-092-154-976
44	27	61.4	88	10	US-09-764-847-976
45	27	61.4	168	10	US-09-925-300-1370

ALIGNMENTS

RESULT 1
US-09-925-301-861
Sequence 861, Application US/09925301
Patent No. US20020052108A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P4106
CURRENT APPLICATION NUMBER: US/09/925,301
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 861
LENGTH: 321
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (18)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-861

Query Match 100.0%; Score 44; DB 10; Length 321;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMFLQIY 9
Db 139 ISEMFLQIY 147
RESULT 2
US-09-996-069-10
Sequence 10, Application US/09996069
Publication No. US20030036199A1
GENERAL INFORMATION:
APPLICANT: Bamdad, Cynthia
APPLICANT: Bamdad, R. Shoshana

```

; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBIT
; FILE REFERENCE: M01015/70071
; CURRENT FILING DATE: 2001-11-27
; CURRENT APPLICATION NUMBER: US/09/996,069
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-996-069-10

Query Match          100.0%; Score 44; DB 9; Length 1255;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
DB 1073 ISEMFLQIY 1081

RESULT 3
US-10-057-136-20
; Sequence 20, Application US/10057336
; Publication No. US20030021770A1
; GENERAL INFORMATION:
; APPLICANT: SCHLOM, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUFE, DONALD
; APPLICANT: PANICALI, DENNIS
; APPLICANT: GRITZ, LINDA
; TITLE OF INVENTION: RECOMBINANT POX VIRUS FOR IMMUNIZATION AGAINST MUC1
; TITLE OF INVENTION: TUMOR-ASSOCIATED ANTIGEN
; FILE REFERENCE: 700953/47113C
; CURRENT APPLICATION NUMBER: US/10/057,136
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 09/366,670
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: PCT/US98/03693
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/038,253
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-057-136-20

Query Match          93.2%; Score 41; DB 9; Length 508;
Best Local Similarity 88.9%; Pred. No. 1.3;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
DB 326 ISEMFLQIY 334

RESULT 4
US-10-087-464-10
; Sequence 10, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishti, Aghar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goei, Yixas
; APPLICANT: Li, Auecong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
```

```

; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; US-10-087-464-10

Query Match          70.5%; Score 31; DB 9; Length 1639;
Best Local Similarity 44.4%; Pred. No. 4.4e+02;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
DB 847 LSQLEWREY 855

RESULT 5
US-09-963-959-7
; Sequence 7, Application US/09963959
; Patent No. US20020165145A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 33521, A NOVEL HUMAN GUANINE NUCLEOTIDE EXCHANGE FAMILY MEMBER
; FILE REFERENCE: 10448-095001
; CURRENT APPLICATION NUMBER: US/09/963,959
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/235,033
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
; US-09-963-959-7

Query Match          68.2%; Score 30; DB 9; Length 207;
Best Local Similarity 55.6%; Pred. No. 89;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
DB 96 LKEPFLQIY 104

RESULT 6
US-10-060-019-27
; Sequence 27, Application US/10060019
; Publication No. US20030003564A1
; GENERAL INFORMATION:
; APPLICANT: Tyers, Mike
; APPLICANT: Willems, Andrew
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
; FILE REFERENCE: 11757.10USU1
; CURRENT APPLICATION NUMBER: US/10/060,019
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: US/09/177,165
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/092,443
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: 60/063,254
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
```

LENGTH: 272
 TYPE: PRT
 ORGANISM: Saccharomyces cerevisiae
 US-10-060-019-27

Query Match 68.2%; Score 30; DB 9; Length 272;
 Best Local Similarity 75.0%; Pred. No. 1.2e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SEMFLQIY 9
 |||||
 Db 98 SEFLQIY 105

RESULT 7
 US-10-060-019-24
 ; Sequence 24, Application US/10060019
 ; Publication No. US20030003564A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tyers, Mike
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
 ; FILE REFERENCE: 11757.10USU1
 ; CURRENT APPLICATION NUMBER: US/10/060,019
 ; PRIOR FILING DATE: 1998-10-22/09/177,165
 ; PRIOR APPLICATION NUMBER: 60/092,443
 ; PRIOR FILING DATE: 1998-07-10
 ; PRIOR APPLICATION NUMBER: 60/063,254
 ; PRIOR FILING DATE: 1997-10-24
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 24
 ; LENGTH: 815
 ; TYPE: PRT
 ; ORGANISM: Saccharomyces cerevisiae
 US-10-060-019-24

Query Match 68.2%; Score 30; DB 9; Length 815;
 Best Local Similarity 75.0%; Pred. No. 3.5e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SEMFLQIY 9
 |||||
 Db 106 SEFLQIY 113

RESULT 8
 US-09-972-970-5
 ; Sequence 5, Application US/09972970
 ; Patent No. US2002016463A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shi et al.
 ; TITLE OF INVENTION: TMA5F Receptor Polynucleotides, Polypeptides, and Antibodies
 ; FILE REFERENCE: PT056P1
 ; CURRENT APPLICATION NUMBER: US/09/972,970
 ; PRIOR FILING DATE: 2001-10-10
 ; PRIOR APPLICATION NUMBER: PCT/US01/11130
 ; PRIOR FILING DATE: 2001-04-05
 ; PRIOR APPLICATION NUMBER: 60/195,336
 ; PRIOR FILING DATE: 2000-04-10
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5
 ; LENGTH: 44
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-972-970-5

Query Match 65.9%; Score 29; DB 9; Length 44;
 Best Local Similarity 44.4%; Pred. No. 30;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ISEMFLQIY 9
 :|:|:|:|:
 Db 25 LSDFLQIY 33

RESULT 9
 US-10-091-504-1005
 ; Sequence 1005, Application US/10091504
 ; Publication No. US2003005908A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC007C1
 ; CURRENT APPLICATION NUMBER: US/10/091,504
 ; CURRENT FILING DATE: 2002-03-07
 ; NUMBER OF SEQ ID NOS: 2442
 ; Prior Application removed - See File Wrapper or Palm
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1005
 ; LENGTH: 69
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (48)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: misc_feature
 ; LOCATION: (49)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: misc_feature
 ; LOCATION: (50)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-10-091-504-1005

Query Match 65.9%; Score 29; DB 9; Length 69;
 Best Local Similarity 62.5%; Pred. No. 48;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SEMFLQIY 9
 |:|:|:|:
 Db 16 SDYFLQIY 23

RESULT 10
 US-09-764-869-1005
 ; Sequence 1005, Application US/09764869
 ; Patent No. US20020061521A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC007
 ; CURRENT APPLICATION NUMBER: US/09/764,869
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - refer to PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 2442
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1005
 ; LENGTH: 69
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (48)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (49)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (50)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-764-869-1005

Query Match 65.9%; Score 29; DB 10; Length 69;
Best Local Similarity 62.5%; Pred. No. 48;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 ISEMFLQIY 9
|:|:|:|
Db 16 SDFIQLY 23

RESULT 11
US-09-925-300-1664
; Sequence 1664, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1664
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1664

Query Match 65.9%; Score 29; DB 10; Length 485;
Best Local Similarity 55.6%; Pred. No. 3.3e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
|:|:|:|
Db 451 IKOMFAOMY 459

RESULT 12
US-10-078-090-209
; Sequence 209, Application US/10078090
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Karra, Kalpana
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenchua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: PRX-0312
; CURRENT APPLICATION NUMBER: US/10/078,090
; PRIOR FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 60/268,999
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 209
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-078-090-209

Query Match 63.6%; Score 28; DB 9; Length 111;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
|:|:|:|
Db 20 ISSIFLSIY 28

RESULT 13
US-10-166-087-34
; Sequence 34, Application US/10166087
; Publication No. US2003007767A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia Biosciences Inc.
; APPLICANT: Farnet, Chris
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: Genes and proteins for the biosynthesis of anthracycin
; FILE REFERENCE: 3014-2US
; CURRENT APPLICATION NUMBER: US/10/166,087
; PRIOR FILING DATE: 2002-06-11
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Streptomyces refuineus subspecies thermotolerans
US-10-166-087-34

Query Match 63.6%; Score 28; DB 9; Length 261;
Best Local Similarity 55.6%; Pred. No. 2.9e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
|:|:|:|
Db 181 IEEVWLEIY 189

RESULT 14
US-09-510-332-1
; Sequence 1, Application US/09510332
; Publication No. US2003002278A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: T2R, a No. US2003002278A1el Family of Taste Receptors
; FILE REFERENCE: 02307E-096010US
; CURRENT APPLICATION NUMBER: US/09/510,332
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 09/393,634
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human T2R01 (hGR01)
US-09-510-332-1

Query Match 63.6%; Score 28; DB 9; Length 299;
Best Local Similarity 44.4%; Pred. No. 3.3e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
|:|:|:|
Db 53 VSRIFLQY 61

RESULT 15
US-09-393-634-35

```

; Sequence 35, Application US/09393634
; Patent No. US20020051997A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: SF, a No. US20020051997A1e1 Family of Taste Receptors
; FILE REFERENCE: 023078-098000US
; CURRENT APPLICATION NUMBER: US/09/393,634
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 299
; TYPE: DRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human GR01, SF01
US-09-393-634-35

Query Match 63.6%; Score 28; DB 10; Length 299;
Best Local Similarity 44.4%; Pred. No. 3.3e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEWFLQIY 9
: * : | | | :
Db 53 VSRIFLQLF 61

Search completed: May 1, 2003, 23:21:44
JOB time : 1.57143 secs

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: May 1, 2003, 23:21:06 ; Search time 15 seconds
(without alignments)
57.681 Million cell updates/sec

Title: US-09-658-621B-26

Perfect score: 44

Sequence: 1 ISEMPLOIY 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 789

Minimum DB seq length: 0

Maximum DB seq length: 9

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	38.6	9	PT0315	Ig heavy chain CRD
2	34.1	6	A27696	contraction-inhibi
3	34.1	7	S08606	hypothetical prote
4	34.1	9	I52974	seminal vesicle pr
5	29.5	9	PT0225	Ig heavy chain CDR
6	29.5	9	I49406	bone gla protein -
7	27.3	4	I54357	schwannomn - mous
8	27.3	6	B27696	contraction-inhibi
9	27.3	7	B39127	phosphotransferase
10	27.3	8	A41117	acetylcholinestera
11	27.3	8	T13818	cytochrome oxidase
12	27.3	9	D58503	transactone elonga
13	27.3	9	B41978	calliperamide 2 -
14	25.0	4	A35779	neuropeptide Antho
15	25.0	5	B45525	actin I - malaria
16	25.0	5	JT0870	phytoestrogen alp
17	25.0	7	NYPG7	hypothalamic hepta
18	25.0	7	B35890	RNA-directed DNA p
19	25.0	7	S09652	hypothetical prote
20	25.0	7	PN0649	pulvinase (EC 3.
21	25.0	8	S66466	cardioscelatory
22	25.0	8	A61467	penalbumin - Adell
23	25.0	8	A35180	neutral proteinase
24	25.0	9	S70334	endosperm protein,
25	25.0	9	S09478	globulin IV, alpha
26	22.7	4	B37325	pap fibroblast regul
27	22.7	5	S70615	endo-1,4-beta-xyla
28	22.7	5	B61445	leu-enkephalin - b

30	10	22.7	6	2	JN0861
31	10	22.7	6	2	B34835
32	10	22.7	6	2	I51317
33	10	22.7	6	2	I49421
34	10	22.7	7	2	S33244
35	10	22.7	7	2	S33245
36	10	22.7	7	2	A58718
37	10	22.7	8	2	B24749
38	10	22.7	8	2	S22428
39	10	22.7	8	2	A46306
40	10	22.7	8	2	A28719
41	10	22.7	8	2	A42057
42	10	22.7	8	2	A25836
43	10	22.7	9	2	B45796
44	10	22.7	9	2	B41466
45	10	22.7	9	2	A92774

ALIGNMENTS

RESULT 1

PT0315
Ig heavy chain CRD3 region (clone 6-109) - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996

C/Accession: PT0315

R/Yamada, M.; Masserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Royers, G.

U. Exp. Med. 173, 395-407, 1991

A/Title: Preferential utilization of specific immunoglobulin heavy chain diversity and

A/Reference number: PT0222; MUID:91108337; PMID:1899102

A/Accession: PT0315

A/Molecule type: DNA

A/Residues: 1-9 <YAM>

A/Keywords: heterotrimer; immunoglobulin

Query Match 38.6%; Score 17; DB 2; Length 9;

Best Local Similarity 42.9%; Pred. No. 2.8e+05;

Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 EMPLOIY 9

DB 2 EXFLSP 8

RESULT 2

A27696
contraction-inhibiting peptide I - blue mussel

C/Species: Mytilus edulis (blue mussel)

C/Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Dec-1995

C/Accession: A27696

R/Hirata, T.; Kubota, I.; Iwasawa, N.; Takabatake, I.; Ikeda, T.; Muneoka, Y.

Biochem. Biophys. Res. Commun. 152, 1376-1382, 1988

A/Title: Structures and actions of Mytilus inhibitory peptides.

A/Reference number: A50142; MUID:86240357; PMID:3377776

A/Accession: A27696

A/Molecule type: protein

A/Residues: 1-6 <HIR>

C/Keywords: amidated carboxyl end (Val) #status experimental

F/6/Modified site: amidated carboxyl end (Val) #status experimental

Query Match 34.1%; Score 15; DB 2; Length 6;

Best Local Similarity 60.0%; Pred. No. 2.8e+05;

Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SEMPL 6

DB 2 SPMV 6

RESULT 3

S08606

Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 MFL 6
|:
Db 4 MFL 6

RESULT 9

B39127 Phosphotransferase system enzyme II (EC 2.7.1.69) - Escherichia coli (fragment)

C/Species: Escherichia coli
C/Date: 27-Nov-1991 #sequence_revision 27-Nov-1991 #text_change 08-Oct-1999

C/Accession: B39127

R/Hardesty, C.; Ferran, C.; Di Rienzo, J.M.

J. Bacteriol. 173, 449-456, 1991

A/Title: Plasmid-mediated sucrose metabolism in Escherichia coli: characterization of sc

rin.

A/Reference number: A39127; MUID:91100329; PMID:1846143

A/Accession: B39127

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-7 <HMR>

A/Cross-references: GB:M38416; NID:g155142; PIDN:AAA98418.1; PID:g155144

C/Keywords: phosphotransferase

Query Match 27.3%; Score 12; DB 2; Length 7;
Best Local Similarity 75.0%; Pred. No. 2.8e+05;
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 FLOI 8
|:
Db 3 FLOI 6

RESULT 10

A41117 acetylcholinesterase (EC 3.1.1.7), venom - Asian cobra (fragment)

C/Species: Naja naja oxiata (Asian cobra, Oxis cobra)

C/Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 23-Jun-1993

C/Accession: A41117

R/Kreienkamp, H.J.; Weise, C.; Raba, R.; Aavikaar, A.; Hucho, F.

Proc. Natl. Acad. Sci. U.S.A. 88, 6117-6121, 1991

A/Title: Anionic subsites of the catalytic center of acetylcholinesterase from Torpedo

A/Reference number: A41117; MUID:91296772; PMID:2068091

A/Accession: A41117

A/Status: preliminary

A/Molecule type: protein

A/Residues: 1-8 <KRB>

C/Keywords: carboxylic ester hydrolase

Query Match 27.3%; Score 12; DB 2; Length 8;
Best Local Similarity 50.0%; Pred. No. 2.8e+05;
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 SEMF 5
|:
Db 2 SEMF 5

RESULT 11

T13818 cytochrome oxidase subunit I - Atlantic hagfish mitochondrion (fragment)

C/Species: mitochondrion Myxine glutinosa (Atlantic hagfish)

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000

C/Accession: T13818

R/Delapierre, C.; Barriel, V.; Tillier, S.; Janvier, P.; Gachelin, G.

Mol. Biol. Evol. 14, 807-813, 1997

A/Title: The main features of the craniate mitochondrial DNA between the NDI and the COI

A/Reference number: Z17775; MUID:97398704; PMID:9254918

A/Accession: T13818

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-8

A/Cross-references: EMBL:Y09527; NID:g2340019; PIDN:CAA0718.1; PID:g2340022

C/Genetics:

A/Genome: mitochondrion

A/Note: COI

C/Keywords: mitochondrion

Query Match 27.3%; Score 12; DB 2; Length 8;
Best Local Similarity 66.7%; Pred. No. 2.8e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 MFL 6
|:
Db 1 MFL 3

RESULT 12

D58503 translation elongation factor EF-Tu - unidentified bacterium (fragment)

C/Species: unidentified bacterium

C/Date: 07-Feb-1997 #sequence_revision 07-Feb-1997 #text_change 28-May-1999

C/Accession: D58503

R/Binette, J.P.; Binette, M.B.

submitted to the Protein Sequence Database, October 1996

A/Description: The proteins of kidney and gallbladder stones.

A/Reference number: A58501

A/Accession: D58503

A/Status: preliminary

A/Molecule type: protein

A/Residues: 1-9 <BIN>

A/Experimental source: human bile and stones

C/Superfamily: translation elongation factor Tu; translation elongation factor Tu homol.

C/Keywords: GTP Binding

Query Match 27.3%; Score 12; DB 2; Length 9;
Best Local Similarity 66.7%; Pred. No. 2.8e+05;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 QTY 9
|:
Db 5 QTY 7

RESULT 13

B41978 Calliophora vomitoria fly (Calliophora vomitoria)

C/Species: Calliophora vomitoria

C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999

C/Accession: B41978

R/Duve, H.; Johnsen, A.H.; Sewell, J.C.; Scott, A.G.; Orchard, I.; Rehfeld, J.F.; Thorp

Proc. Natl. Acad. Sci. U.S.A. 89, 2326-2330, 1992

A/Title: Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH-2 neuropeptides (des)

A/Reference number: A41978; MUID:92196111; PMID:1549595

A/Accession: B41978

A/Status: preliminary

A/Molecule type: protein

A/Residues: 1-9 <DUV>

C/Keywords: amidated carboxyl end; neuropeptide

F/9/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 27.3%; Score 12; DB 2; Length 9;
Best Local Similarity 33.3%; Pred. No. 2.8e+05;
Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SEMFLO 7
|:
Db 3 SEMFLO 8

RESULT 14

C41978 Calliophora vomitoria fly (Calliophora vomitoria)

C/Species: Calliophora vomitoria

C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999

C:/Accession: C41978
 R:/Dive, H.; Johnsen, A.H.; Sewell, J.C.; Scott, A.G.; Orchard, I.; Rehfeld, J.F.; Thorpe
 Proc. Natl. Acad. Sci. U.S.A. 89, 2326-2330, 1992
 A:/Title: Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH-2 neuropeptides (desi
 A:/Reference number: A41978; MUID:92196111; PMID:1549595
 A:/Accession: C41978
 A:/Status: preliminary
 A:/Molecule type: protein
 A:/Residues: 1-9 <DUV>
 C:/Keywords: amidated carboxyl end; neuropeptide
 F:/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 27.3%; Score 12; DB 2; Length 9;
 Best Local Similarity 33.3%; Pred. No. 2.8e+05;
 Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 SEMFLO 7
 | : | : | :
 DB 3 SODPMK 8

RESULT 15

A35779
 neuropeptide Antho-RNamide - sea anemone (Anthopleura elegantissima)
 C:/Species: Anthopleura elegantissima
 C:/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 08-Dec-1995
 C:/Accession: A35779
 R:/Grimmelikhuizen, C.J.P.; Rinehart, K.L.; Jacob, E.; Graff, D.; Reinscheid, R.K.; Noth
 Proc. Natl. Acad. Sci. U.S.A. 87, 5410-5414, 1990
 A:/Title: Isolation of L-3-phenyllactyl-Leu-Arg-Asn-NH2 (Antho-RNamide), a sea anemone ne
 A:/Reference number: A35779; MUID:90319122; PMID:1973541
 A:/Accession: A35779
 A:/Molecule type: protein
 A:/Residues: 1-4 <GR1>
 C:/Comment: The L-3-phenyllactyl blocking group probably arises from an amino-terminal ph
 C:/Keywords: amidated carboxyl end; neuropeptide; phenyllactylation
 F:/Modified site: L-3-phenyllactic acid (Phe) #status experimental
 F:/Modified site: amidated carboxyl end (Asn) #status experimental

Query Match 25.0%; Score 11; DB 2; Length 4;
 Best Local Similarity 66.7%; Pred. No. 2.8e+05;
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 FLQ 7
 | : | : | :
 DB 1 FLR 3

Search completed: May 1, 2003, 23:24:24
 Job time : 16 secs

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 13:22:52 ; Search time 396 Seconds

(without alignments)
8939,751 Million cell updates/sec

Title: US-09-658-621B-1

Perfect score: 1572
Sequence: 1 gaattccctgcgcttgtaa.....tcgccttcgcgctgagtcg 1572

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq.101002.*
1: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1572	100.0	1572	22	AA500585 Human MUC1 DNA seq
2	1481.2	94.2	1721	24	ABL68032 Ovary cancer relat
3	1481.2	94.2	1721	24	ABL68032 Human ovarian tumor
4	1412.2	89.8	1804	24	ABL67539 Thyroid cancer relat
5	1342	85.4	1428	24	ABL60159 Human MUC1 encodin
6	1313	83.5	1452	13	AAQ29276 Encodes transmembr
7	1257	80.0	1527	19	AAV48329 MIM10C1 gene. Ho
8	1213.4	77.2	1737	21	UBIQUITIN-E. coli
9	1210.2	77.0	1371	21	AA000388 Human Mucin 1 (MUC

10	1190.2	75.7	4139	24	ABN95623	Gene #2121 used to
11	1190.2	75.7	4139	24	ABE7071	Thyroid cancer relat
12	1190.2	75.7	4139	24	ABE7534	Thyroid cancer relat
13	1190.2	75.7	4139	24	ABE6884	Kidney cancer rela
14	1190.2	75.7	4139	24	ABK09797	Human ovarian tumor
15	1056	67.2	1074	13	AAQ24679	H23-ETa-T antigen
16	993.4	62.2	1320	16	AAI29190	Glycoprotein 39.3
17	978	62.2	1355	21	AACT7625	Human cancer assoc
18	750	47.7	6286	24	ABE60158	Human MUC1 encodin
19	750	47.7	8181	24	ABE41001	Human neurogulin 2
20	738.4	47.0	1709	13	AAQ29277	Encodes secreted f
21	661.4	42.1	822	17	AAI10677	Mucin-derived prot
22	661.4	42.1	849	17	AAI10678	Mucin-derived prot
23	638.2	40.6	1020	21	AAQ00392	Ubiquitin-E. coli
24	635.4	40.6	654	21	AAQ00386	Human Mucin 1 (MUC
25	607.4	38.6	768	17	AAI10679	Mucin-derived prot
26	607.4	38.6	795	17	AAI10680	Mucin-derived prot
27	587.2	37.4	723	17	AAI10681	Mucin-derived prot
28	587.2	37.4	750	17	AAI10682	H23-ETa-T antigen
29	540	34.4	558	13	AAQ24678	H23-ETa-S antigen
30	494.4	31.5	497	24	ABK09657	Human ovarian tumor
31	481.4	30.6	1331	13	AAQ24681	H23-ETa-S antigen
32	459	29.2	5660	24	ABK34357	Human immune syste
33	429.2	27.3	459	12	AAQ12772	Human polymorphic
34	429.2	27.3	891	21	AAQ00391	Ubiquitin-E. coli
35	366	23.3	891	21	AAQ00385	Human Mucin 1 (MUC
36	362.8	23.1	525	21	AAQ00385	Human Mucin 1 (MUC
37	328.8	20.9	1763	12	AAQ12771	Human lung tumor
38	328.4	20.9	330	22	AAE68091	Human lung tumor
39	328.4	20.9	330	22	ABK8002	CDNA encoding clon
40	285.8	18.2	1567	12	AAQ12773	Human polymorphic
41	279	17.7	285	21	AAQ00387	Human Mucin 1 (MUC
42	277	17.6	309	10	AAQ00393	Ubiquitin-E. coli
43	253.6	16.1	309	10	AAQ00393	PDPA.3 CDNA insert
44	237	15.1	237	24	ABK09556	Human ovarian tumor
45	235.4	15.0	5660	24	ABK34356	Human immune syste

ALIGNMENTS

RESULT 1	AA500585	standard; DNA; 1572 BP.
ID	AA500585	
AC	AA500585	
XX	12-SEP-2001	(first entry)
DT		
XX		
DE	Human MUC1 DNA sequence.	
XX		
KW	Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-I;	
KW	glycoprotein; cytotoxic T lymphocytes; T cell response; cancer; vaccine;	
KW	cancer gene therapy; diagnosis; treatment; inflammatory disorder; de;	
KW	organ transplant rejection; graft versus host disease.	
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	58..1545
FT		/+tag= a
FT		/product= "Human MUC1"
XX		
XX	MO200118035-A2.	
XX		
PD	15-MAR-2001.	
XX		
PF	07-SEP-2000; 2000MO-EP08761.	
XX		
PR	08-SEP-1999; 99GB-0021242.	
PR	10-SEP-1999; 99EP-040237.	
PR	03-MAR-2000; 2000US-0187215.	
XX		

PA (TRGE) TRANSGENE SA.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
PI Taylor-Papadimitriou J, Heukamp LC, Offringa R, Melief CJM;
PI Acres B, Thomas M;
XX
XX MPI: 2001-235187/24.
DR P-PSDB; AAU00539.
XX
XX
PT New antigenic polypeptides of MHC-I protein which activate cytotoxic T
PT lymphocyte proteins and their analogues, useful for identifying a major
PT histocompatibility complex class I restricted T cell response and for
PT diagnosing cancer -
XX
XX Claim 1; Page 66-68; 81pp; English.
XX
XX The sequence represents a polynucleotide which encodes MHC1 polypeptide.
CC derivative antigenic peptides of this protein bind at least one major
CC histocompatibility complex class I (MHC-I) glycoprotein, which activates
CC cytotoxic T lymphocytes to induce a protective response against tumours.
CC Diagnosis of cancer involves determining the presence or absence in a
CC host cell of a MHC class I restricted T cell response to a MHC1
CC derivative, where the presence of the MHC class I restricted T cell
CC response indicates that the host has cancer. Measurement of the level of
CC MHC class I restricted T cell response is also useful to monitor the
CC severity of cancer, a larger response indicating a more severe cancer.
CC MHC1 derivatives are useful in cancer therapy and to follow MHC1 specific
CC immune responses in patients during the course of disease and/or
CC treatment. MHC1 DNA is useful in cancer gene therapy, vaccination and
CC diagnosis. Compositions of the sequences are used in vaccines and
CC as an inflammatory disorder, organ transplant rejection or graft versus
CC host disease.
XX
XX Sequence 1572 BP; 312 A; 567 C; 364 G; 329 T; 0 other;
SQ
Query Match 100.0%; Score 1572; DB 22; Length 1572;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCCTGCTGCTTGAATCTGTTCTGCCCCCTCCCAACCCATTTTCAACCAACCATG 60
DB 1 GAATTCCTGCTGCTTGAATCTGTTCTGCCCCCTCCCAACCCATTTTCAACCAACCATG 60
QY 61 AACCGGGGACACCGATGTCCTTCTTCTGCTGCTGCTCTCAAGTCTTACAGTGT 120
DB 61 AACCGGGGACACCGATGTCCTTCTTCTGCTGCTGCTCTCAAGTCTTACAGTGT 120
QY 121 ACAGGTTTGTGATGCAAGCTTACCCAGGTGAGAAAGAGATTGGCTACCCAG 180
DB 121 ACAGGTTTGTGATGCAAGCTTACCCAGGTGAGAAAGAGATTGGCTACCCAG 180
QY 181 AGAAGTTCAAGTCCCAAGCTCTAATGAAAGAAATCTGTGATATGACAGAGGCTACT 240
DB 181 AGAAGTTCAAGTCCCAAGCTCTAATGAAAGAAATCTGTGATATGACAGAGGCTACT 240
QY 241 TCCAGCCACAGACCCCGGTTCAAGCTCTCTCAACCACTAGAGGACAGATGTCACTGGCC 300
DB 241 TCCAGCCACAGACCCCGGTTCAAGCTCTCTCAACCACTAGAGGACAGATGTCACTGGCC 300
QY 301 CGGCGCAAGGAAACAGATTCAAGTTCAAGTCCCACTTGGGAGACAGATGTCACTGGCT 360
DB 301 CGGCGCAAGGAAACAGATTCAAGTTCAAGTCCCACTTGGGAGACAGATGTCACTGGCT 360
QY 361 CGAGTCAACAGGACCGCTTGAGCTTCAACCAACCCCGCAGCCCAAGATGTCACTGGCC 420
DB 361 CGAGTCAACAGGACCGCTTGAGCTTCAACCAACCCCGCAGCCCAAGATGTCACTGGCC 420
QY 421 CCGGACAAAGAGCCCGGAGCTTCAACCGCCCGCCGAGGATGTCACTGGCC 480
DB 421 CCGGACAAAGAGCCCGGAGCTTCAACCGCCCGCCGAGGATGTCACTGGCC 480
QY 481 CCGGACAAAGAGCCCGGAGCTTCAACCGCCCGCCGAGGATGTCACTGGCC 540

DB 481 CCGGACAAAGAGCCCGGAGCTTCAACCGCCCGCCGAGGATGTCACTGGCC 540
QY 541 CCGGACAAAGAGCCCGGAGCTTCAACCGCCCGCCGAGGATGTCACTGGCC 600
DB 541 CCGGACAAAGAGCCCGGAGCTTCAACCGCCCGCCGAGGATGTCACTGGCC 600
QY 601 CCGGACAAAGAGCCCGGAGCTTCAACCGCCCGCCGAGGATGTCACTGGCC 660
DB 601 CCGGACAAAGAGCCCGGAGCTTCAACCGCCCGCCGAGGATGTCACTGGCC 660
QY 661 TCAAGGCTTGAATGAGGCTTCAAGTCTTCACTGTGTGCAACAGGACCTTGGCAAGGCT 720
DB 661 TCAAGGCTTGAATGAGGCTTCAAGTCTTCACTGTGTGCAACAGGACCTTGGCAAGGCT 720
QY 721 ACCACAAACCCAGCCAGACCAAGACCTCCACCCAGCAATTCAGCAGCACTGTGATCT 780
DB 721 ACCACAAACCCAGCCAGACCAAGACCTCCACCCAGCAATTCAGCAGCACTGTGATCT 780
QY 781 CTTACACACCTTGGCAGCAGTACAGACCAAGATGTGATGATGATGATGATGATGATGATG 840
DB 781 CTTACACACCTTGGCAGCAGTACAGACCAAGATGTGATGATGATGATGATGATGATGATG 840
QY 841 GTACCTCTCTCACTCTCTCAATTCACAGACCTTCTCCAGTGTCTTACCTGGGCTCT 900
DB 841 GTACCTCTCTCACTCTCTCAATTCACAGACCTTCTCCAGTGTCTTACCTGGGCTCT 900
QY 901 TTTCTTTTCTGCTTTTCAATTTCAACCTCCAGTTTAAATCTCTCTGAGAGATCC 960
DB 901 TTTCTTTTCTGCTTTTCAATTTCAACCTCCAGTTTAAATCTCTCTGAGAGATCC 960
QY 961 AGACGCACTCTCAAGAGCTGCAAGAGACATTTCTGAATGTTTTTGGATTTAT 1020
DB 961 AGACGCACTCTCAAGAGCTGCAAGAGACATTTCTGAATGTTTTTGGATTTAT 1020
QY 1021 AAACAAAGGGGTTTTTCTGGGCTCTTCCAAATTAATTAAGTTCAAGCTCAGATCTGTGTGTA 1080
DB 1021 AAACAAAGGGGTTTTTCTGGGCTCTTCCAAATTAATTAAGTTCAAGCTCAGATCTGTGTGTA 1080
QY 1081 CAATTAAGCTCAGGCTTCCGAGAGAGTACATCAATGTCCACGACCTGGAGACACAGTTTC 1140
DB 1081 CAATTAAGCTCAGGCTTCCGAGAGAGTACATCAATGTCCACGACCTGGAGACACAGTTTC 1140
QY 1141 AATCAGTATTAAGAGAGAGAGCTCTGATATTAACCTGACATCTCAGAGCTCAGGCTG 1200
DB 1141 AATCAGTATTAAGAGAGAGAGCTCTGATATTAACCTGACATCTCAGAGCTCAGGCTG 1200
QY 1201 AGTCATGTGCAATTTCTTTCTGTCGCAAGTCTGGGCTGGGAGTCCAGGCTGGGCAATC 1260
DB 1201 AGTCATGTGCAATTTCTTTCTGTCGCAAGTCTGGGCTGGGAGTCCAGGCTGGGCAATC 1260
QY 1261 GCGCTGTGCTGCTGATCTGTGTTCTGTTGCGGTGCGCAATGTTCTATCTCAATTCCTTG 1320
DB 1261 GCGCTGTGCTGCTGATCTGTGTTCTGTTGCGGTGCGCAATGTTCTATCTCAATTCCTTG 1320
QY 1321 GCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
DB 1321 GCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
QY 1381 ACCTAATCTCTTAAGAGAGAGTACCAACCTTACCAACCAATGGGAGCTATGCTGCTGCT 1440
DB 1381 ACCTAATCTCTTAAGAGAGAGTACCAACCTTACCAACCAATGGGAGCTATGCTGCTGCT 1440
QY 1441 AGCAGTACCAATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
DB 1441 AGCAGTACCAATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
QY 1501 TCTTACAAACCAAGAGAGTGGAGGCACTTCTGCAACTTGTAGGGGAGAGTGGCTGCTGCT 1560
DB 1501 TCTTACAAACCAAGAGAGTGGAGGCACTTCTGCAACTTGTAGGGGAGAGTGGCTGCTGCT 1560
QY 1561 TGAAGTGAATGG 1572

[illegible]

	RESULT 3
ID	ABK09743 standard; cDNA; 1721 BP.
AC	ABK09743;
XX	
DT	14-MAR-2002 (first entry)
DE	Human ovarian tumour protein encoding cDNA #276.
XX	
KW	Human; ovarian tumour protein; cancer; cytostatic; immunostimulant; ss; gene therapy; CD4+ T cell; CD8+ T cell; PCR primer.
OS	Homoe sapiens.
XX	
FA	MO200190154-A2.
FN	
ED	29-NOV-2001.
PD	
PF	23-MAY-2001; 2001WO-US16895.
PP	
PX	14-MAY-2000; 2000US-207107P.
PR	13-JUN-2000; 2000US-211457P.
PR	21-JUN-2000; 2000US-213673P.
PR	03-AUG-2000; 2000US-223288P.
PR	01-MAR-2001; 2001US-272750P.
PA	(CORI-) CORIXA CORP.
XX	
XX	Xu J, Mitcham JL, Harlocker SL, Dillon DC, Secrist H, Lodes MJ; Algate PA, Fling SP, Mannon J, Benson DR, Carter D; MPI; 2002-097641/13.
DR	
PT	New isolated polynucleotide encoding polypeptide comprising portion of ovarian tumour protein, useful for detection, diagnosis and therapy of human ovarian cancer -
XX	
PS	Claim 1; Page 222; 285pp; English.
CC	The invention relates to an isolated polynucleotide encoding a polypeptide comprising a portion of an ovarian tumour protein. The sequences of the invention are useful for stimulating an immune response and for treating ovarian cancer in a patient. An antigen presenting cell that expresses the sequences is useful for treating ovarian cancer by incubating CD4+ and/or CD8+ T cells isolated from a patient. The T cells can then be proliferated and administered to the patient to inhibit the development of cancer. The DNA sequences are useful as probes or primers for nucleic acid hybridization, to direct expression of a polypeptide in appropriate host cells. Detecting the presence of a cancer in a patient involves obtaining a biological sample from the patient, detecting the amount of protein that binds to the agent, comparing the amount of protein to a predetermined cut-off value and determining the presence of cancer. Sequences ABK09464-ABK09802 represent PCR primers and cDNA molecules encoding ovarian tumour proteins of the invention.
XX	
SQ	Sequence 1721 BP; 339 A; 634 C; 395 G; 353 T; 0 other;
Query March	94.2%; Score 1481.2; DB 24; Length 1721;
Best Local Similarity	95.5%; Pseq No 1.6e-311;
Matches 1559;	Conservative 0; Mismatches 13; Indels 60; Gaps 1
OY	1 GAATTCCTGGGCGTGAATCTGTTCCTGCCGCCTCCCAAGATTACACACACACTG
DG	1 CAATTTCTGGCTTGAAATCATTCTTCCCTCCCAACCATTTACACACACACTG
OY	61 ACACCGGACACCAGTCTCTTCTTCTGCTGCTGCTCCACAGTGTACAGTGTGTT
DG	61 ACACCGGACACCAGTCTCTTCTTCTGCTGCTGCTCCACAGTGTACAGTGTGTT
OY	121 ACAAGTTCTGTGATGACAAGCTTCAACCCAGATGAGAAAAGAGAAGACCTTGGCTACCGAG
DG	121 ACAAGTTCTGTGATGACAAGCTTCAACCCAGATGAGAAAAGAGAAGACCTTGGCTACCGAG


```
Db 121 ACAGGTTCTGTCATGCAAGCTCTAACCCAGGTGAGAAAGAGAGACTTCGGCTACCCAG 180
Qy 181 AGAAGTTCAATGCCCCAGCTCTACTGAGAGAAAGTGTGATGATGACAGCAGCTACTC 240
Db 181 AGAAGTTCAATGCCCCAGCTCTACTGAGAGAAAGTGTGATGATGACAGCAGCTACTC 240
Qy 241 TCCAGCCAGAGCCCCGGGTTCAAGGCTCTCCAGCACTCAGGGACAGAGATGTCACCTGGCC 300
Db 241 TCCAGCCAGAGCCCCGGGTTCAAGGCTCTCCAGCACTCAGGGACAGAGATGTCACCTGGCC 300
Qy 301 CCGGCAACGAGAACCAAGCTTCAAGTTCAAGTCCACTGGGGAGACAGATGTCACCTGGCTC 360
Db 301 CCGGCAACGAGAACCAAGCTTCAAGTTCAAGTCCACTGGGGAGACAGATGTCACCTGGCTC 360
Qy 361 CCAAGTCAACAGGCAAGCCCTGAGGCTTCACCAAGCCCGGCAAGCCAGATGTCACCTGAGCC 420
Db 361 CCAAGTCAACAGGCAAGCCCTGAGGCTTCACCAAGCCCGGCAAGCCAGATGTCACCTGAGCC 420
Qy 421 CCGGCAACAAAGCAAGCCCGGAGCTCCACGAGCCCGGCGGCAAGATGTCACCTGGAGCC 480
Db 421 CCGGCAACAAAGCAAGCCCGGAGCTCCACGAGCCCGGCGGCAAGATGTCACCTGGAGCC 480
Qy 481 CCGGCAACAGAGGCGGCGGCGGAGCTCCACGAGCCCGGCGGCAAGATGTCACCTGGAGCC 513
Db 481 CCGGCAACAGAGGCGGCGGCGGAGCTCCACGAGCCCGGCGGCAAGATGTCACCTGGAGCC 513
Qy 514 -----CCGCGCGCCACAGTGTCACTCGGCC 540
Db 514 CCGGCAACAGAGGCGGCGGCGGAGCTCCACGAGCCCGGCAAGCCAGATGTCACCTGGAGCC 540
Qy 541 CCGGCAACAGAGGCGGCGGCGGAGCTCCACGAGCCCGGCGGCAAGATGTCACCTGGAGCC 600
Db 541 CCGGCAACAGAGGCGGCGGCGGAGCTCCACGAGCCCGGCGGCAAGATGTCACCTGGAGCC 600
Qy 601 CCGGCAACAGAGGCGGCGGCGGAGCTCCACGAGCCCGGCGGCAAGATGTCACCTGGAGCC 660
Db 601 CCGGCAACAGAGGCGGCGGCGGAGCTCCACGAGCCCGGCGGCAAGATGTCACCTGGAGCC 660
Qy 661 CCGGCAACAGAGGCGGCGGCGGAGCTCCACGAGCCCGGCGGCAAGATGTCACCTGGAGCC 720
Db 661 CCGGCAACAGAGGCGGCGGCGGAGCTCCACGAGCCCGGCGGCAAGATGTCACCTGGAGCC 720
Qy 721 TCAAGCTCTGATTCAGATCAGCTCTACTGATGTCAGAGAGCAAGCTTCGAGAGGCT 780
Db 721 TCAAGCTCTGATTCAGATCAGCTCTACTGATGTCAGAGAGCAAGCTTCGAGAGGCT 780
Qy 781 ACCCAACACCCAGCAGCAAGAGCACTCATTCTCAATTCAGAGCAACCACTCTGATACT 840
Db 781 ACCCAACACCCAGCAGCAAGAGCACTCATTCTCAATTCAGAGCAACCACTCTGATACT 840
Qy 841 CCTACCAACCCCTTGCCAGCATAGCAAGCAAGTATGTCAGATGTCACCATAGAGCA 900
Db 841 CCTACCAACCCCTTGCCAGCATAGCAAGCAAGTATGTCAGATGTCACCATAGAGCA 900
Qy 901 GTACCTCTCTCACTCTCTCAATCAAGCACTTCTCCAGATGTCATCTGGAGGCTCT 960
Db 901 GTACCTCTCTCACTCTCTCAATCAAGCACTTCTCCAGATGTCATCTGGAGGCTCT 960
Qy 961 TTTCTTTTCTGTCGTTTTCATATTCAAACTCCAGTTTAACTTCTCTCTGAAAGATCCC 1020
Db 961 TTTCTTTTCTGTCGTTTTCATATTCAAACTCCAGTTTAACTTCTCTCTGAAAGATCCC 1020
Qy 1021 AGCAGCAGCTACTACCAAGAGCTGCAAGAGACATTTCTGAATGCTTTTTCAGATTTAT 1080
Db 1021 AGCAGCAGCTACTACCAAGAGCTGCAAGAGACATTTCTGAATGCTTTTTCAGATTTAT 1080
Qy 1081 AAAAGAGGGGGTTTTTCTGGGCTCTCCAAATTTAAAGTTTCAAGGCGAGATCTGTGGTAT 1140
Db 1081 AAAAGAGGGGGTTTTTCTGGGCTCTCCAAATTTAAAGTTTCAAGGCGAGATCTGTGGTAT 1140
Qy 1081 CAATTGATCTGAGCTTCGAGAGAGTACATCAATGTCACAGAGAGAGACAGATTC 1200
Db 1141 CAATTGATCTGAGCTTCGAGAGAGTACATCAATGTCACAGAGAGAGACAGATTC 1200
Qy 1141 AATCAGTATAAAGCAAGAGAGCTTCGATATATACTGACATTCAGAGCTGAGCGTG 1200
Db 1201 AATCAGTATAAAGCAAGAGAGCTTCGATATATACTGACATTCAGAGCTGAGCGTG 1260
```

```
Qy 1201 AGTATATGACCATTTCTTTCTCTGCCCAAGTCTGGGAGCTGGAGTCCAGGCTGGAGCATC 1260
Db 1261 AGTATATGACCATTTCTTTCTCTGCCCAAGTCTGGGAGCTGGAGTCCAGGCTGGAGCATC 1320
Qy 1261 GCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320
Db 1321 GCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
Qy 1321 GCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
Db 1381 GCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1440
Qy 1381 ACCTACATCTTATGAGAGATACCCCACTTACCAACCAATGAGGCGCTATGTCCTCCCT 1440
Db 1441 ACCTACATCTTATGAGAGATACCCCACTTACCAACCAATGAGGCGCTATGTCCTCCCT 1500
Qy 1441 AGCAGTACCGATCTGAGCCCTTATGAGAGAGTTCGAGAGTAAATGATGAGAGAGCTTC 1500
Db 1501 AGCAGTACCGATCTGAGCCCTTATGAGAGAGTTCGAGAGTAAATGATGAGAGAGCTTC 1560
Qy 1501 TCTTACAAACCCAGAGAGGAGAGCACTTCTGCAACTTGTAGAGGAGACGTGCCCTC 1560
Db 1561 TCTTACAAACCCAGAGAGGAGAGCACTTCTGCAACTTGTAGAGGAGACGTGCCCTC 1620
Qy 1561 TGAAGCTGAGTGG 1572
Db 1621 TGAAGCTGAGTGG 1632

RESULT 4
ABL67539
ID ABL67539 standard; DNA; 1804 BP.
XX
AC ABL67539;
XX
DT 15-MAY-2002 (first entry)
XX
DE Thyroid cancer related gene sequence SEQ ID NO:5876.
XX
KW Human; cancer; colon; breast; ovary; esophagus; kidney; thyroid;
KM stomach; lung; prostate; pancreas; carcinoma; melanoma; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilms tumor; adenocarcinoma;
gene; db.
XX
OS Homo sapiens.
XX
PN M0200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001MC-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
XX
PR 18-SEP-2000; 2000US-223133P.
XX
PR 18-SEP-2000; 2000US-233617P.
XX
PR 20-SEP-2000; 2000US-234009P.
XX
PR 20-SEP-2000; 2000US-234052P.
XX
PR 22-SEP-2000; 2000US-234509P.
XX
PR 22-SEP-2000; 2000US-234567P.
XX
PR 25-SEP-2000; 2000US-234923P.
XX
PR 25-SEP-2000; 2000US-234924P.
XX
PR 25-SEP-2000; 2000US-235077P.
XX
PR 25-SEP-2000; 2000US-235082P.
XX
PR 25-SEP-2000; 2000US-235134P.
XX
PR 25-SEP-2000; 2000US-235280P.
XX
PR 26-SEP-2000; 2000US-235637P.
XX
PR 26-SEP-2000; 2000US-235638P.
XX
PR 27-SEP-2000; 2000US-235711P.
XX
PR 27-SEP-2000; 2000US-235720P.
XX
PR 27-SEP-2000; 2000US-235840P.
```


QY 838 AGGTAACCTCTCACTCTCAATCAAGACATTCTCCCAATTGTTCTACTGAGGATC 897
 DB 721 AGGGTACTCTCTCACTCTCTCAATCAAGACATTCTCCCAATTGTTCTACTGAGGATC 780
 QY 898 TCTTTCTTTTCTCTGCTCTTTCATTTTCAAACTCCAGATTATTTCTCTGGAAGAT 957
 DB 781 TCTTTCTTTTCTCTGCTCTTTCATTTTCAAACTCCAGATTATTTCTCTGGAAGAT 840
 QY 958 CCGAGCACCGACATCAACCAAGAGCTGACAGACATTTTGAAATGTTTTCAGATT 1017
 DB 841 CCGAGCACCGACATCAACCAAGAGCTGACAGACATTTTGAAATGTTTTCAGATT 900
 QY 1018 TATTAACAAAGGAGGATTTCTGAGGCTCTCAATATTAAGTTGAGGCGAGATCTGAGTG 1077
 DB 901 TATTAACAAAGGAGGATTTCTGAGGCTCTCAATATTAAGTTGAGGCGAGATCTGAGTG 960
 QY 1078 GATCAATTTGACTGAGGCTCTGAGAGAGGTAACATCAATGTTCAAGAGTGGAGACAG 1137
 DB 961 GATCAATTTGACTGAGGCTCTGAGAGAGGTAACATCAATGTTCAAGAGTGGAGACAG 1020
 QY 1138 TTCAATCAATTAATAACGAGACAGCTCTGATATTAACCTGATCGATCTGACAGCTGAC 1197
 DB 1021 TTCAATCAATTAATAACGAGACAGCTCTGATATTAACCTGATCGATCTGACAGCTGAC 1080
 QY 1198 GTGAGTCAATGTCATTTCTTTCTCTGCGCAGTCTGAGGCTGAGGCTGAGGCTGAGGCT 1257
 DB 1081 GTGAGTCAATGTCATTTCTTTCTCTGCGCAGTCTGAGGCTGAGGCTGAGGCTGAGGCT 1140
 QY 1258 ATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
 DB 1141 ATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
 QY 1318 TTGCT 1377
 DB 1201 TTGCT 1260
 QY 1378 GATACCTACCATCTTATGAGCAGATGCCACCTACCAACCATGAGGCTGATGAGGCC 1437
 DB 1261 GATACCTACCATCTTATGAGCAGATGCCACCTACCAACCATGAGGCTGATGAGGCC 1320
 QY 1438 CCGAGCACCGACATCAACCAAGAGCTGACAGACATTTTGAAATGTTTTCAGATT 1497
 DB 1321 CCGAGCACCGACATCAACCAAGAGCTGACAGACATTTTGAAATGTTTTCAGATT 1380
 QY 1498 CCGAGCACCGACATCAACCAAGAGCTGACAGACATTTTGAAATGTTTTCAGATT 1545
 DB 1381 CCGAGCACCGACATCAACCAAGAGCTGACAGACATTTTGAAATGTTTTCAGATT 1428

RESULT 6
 AAQ29276 ID AAQ29276 standard; cDNA; 1452 BP.
 AC AAQ29276;
 DT 06-NOV-1992 (first entry)
 XX
 DE Encodes transmembrane form of H23-ETA antigen.
 XX
 KM ETA-T; human epithelial antigen; Monoclonal antibody H23; vaccine;
 KM malignant tumour; breast cancer; ss.
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT sig_peptide /tag= a
 FT mat_peptide /tag= b
 FT misc_feature /tag= c
 FT /note= "epitope recognised by Mab H23, can be present in 1-80 tandem repeats"

FT variation 457..459
 FT /tag= d
 FT /note= "either an Ala or a Pro codon"
 FT variation 487..489
 FT /tag= e
 FT /note= "either a Thr or an Asn codon"
 FT variation 496..498
 FT /tag= f
 FT /note= "either a Ala or an Pro codon"
 PN MO9207000-A.
 XX 30-APR-1992.
 XX
 PF 23-OCT-1991; 91WO-FR00835.
 XX
 PR 23-OCT-1990; 90FR-0013101.
 XX
 DR (TRGE) TRANSGENE SA.
 XX
 PI Chamhon P, Hareuveni M, Kieny MP, Lathe R,
 XX
 DR WPI; 1992-167097/20.
 XX
 PT P-PSDB; AAR23973.
 PT Compens. conty. polypeptide antigen recognised by antibody H23 -
 PT for treatment of mammary tumours, also for pox virus compens. for
 PT use in vaccines
 XX
 PS Claim 3; Page 15-18; 29pp; French.
 XX
 CC The tumour recognised by antibody H23 is aberrantly
 CC expressed in epithelial cells from cancerous mammary tissue in about
 CC 90 per cent of breast cancer cases; in a normal individual
 CC expression is negligible. The antigen exists in two forms:
 CC transmembrane (ETA-T) and secreted (ETA-S). Both forms show a high
 CC degree of polymorphism. A 20 amino acid subunit in ETA can be
 CC tandemly repeated up to 80 times. From one subunit to the next, 1 to
 CC 3 amino acids can differ (= "variation" in the features table). DNA
 CC coding for immunogenic fragments of ETA can be inserted into
 CC e.g. vaccinia viruses for treatment of mammary tumours.
 CC See also AAQ24678-Q24681, AAQ29277 and AAR23974-R23981.
 XX

SO Sequence 1452 BP; 298 A; 495 C; 329 G; 321 T; 9 other:
 Query Match 83.5%; Score 1313; DB 13; Length 1452;
 Best Local Similarity 91.8%; Pred. No. 4.4e-275;
 Matches 1443; Conservative 0; Mismatches 9; Indels 120; Gaps 1;
 QY 1 GAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
 DB 1 GAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
 QY 61 ACACCGGACACCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
 DB 61 ACACCGGACACCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
 QY 121 ACAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
 DB 121 ACAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
 QY 181 AGAAGTTCAAGTCCCACTCTACTGAGAAAGATCTGTGATGATGACGAGGACTCTC 240
 DB 181 AGAAGTTCAAGTCCCACTCTACTGAGAAAGATCTGTGATGATGACGAGGACTCTC 240
 QY 241 TCCAGCAGACGCCCCGGTTCAAGCTCTCCACCACTGAGGAGACAGATGACTCTGAGCC 300
 DB 241 TCCAGCAGACGCCCCGGTTCAAGCTCTCCACCACTGAGGAGACAGATGACTCTGAGCC 300
 QY 301 CCGGCCACGGAACCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCA 360
 DB 301 CCGGCCACGGAACCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCA 360

```

Oy 361 CGAGTACCAAGGCCAGCCCTGAGGCTCCACCAACCCCGCCAGCCCAAGATGTCACCTCAGCC 420
Db 361 CGAGTACCAAGGCCAGCCCTGAGGCTCCACCAACCCCGCCAGCCCAAGATGTCACCTCAGCC 420
Oy 421 CCGGACAAACAAAGCCAGCCCGGAGCTCCAGCCGCGCCCGGAGGCTGTCACCTCGGACC 480
Db 421 CCGGACAAACAAAGCCAGCCCGGAGCTCCAGCCGCGCCCGGAGGCTGTCACCTCGGACC 480
Oy 481 CCGGACAAACAAAGCCAGCCCGGAGCTCCAGCCGCGCCCGGAGGCTGTCACCTCGGACC 540
Db 481 CCGGACAAACAAAGCCAGCCCGGAGCTCCAGCCGCGCCCGGAGGCTGTCACCTCGGACC 540
Oy 541 CCGGACAAACAAAGCCAGCCCGGAGCTCCAGCCGCGCCCGGAGGCTGTCACCTCGGACC 600
Db 541 CCGGACAAACAAAGCCAGCCCGGAGCTCCAGCCGCGCCCGGAGGCTGTCACCTCGGACC 600
Oy 489 ----- 488
Db 489 ----- 488
Oy 601 CCGGACAAACAAAGCCAGCCCGGAGCTCCAGCCGCGCCCGGAGGCTGTCACCTCGGACC 660
Db 489 ----- 540
Oy 661 TCAGGCTGTCAGTCAAGGCTCAGCTTCTAATCTGTCAGTCAAAAGGCACTCTGCGAGGGCT 720
Db 541 TCAGGCTGTCAGTCAAGGCTCAGCTTCTAATCTGTCAGTCAAAAGGCACTCTGCGAGGGCT 600
Oy 721 ACCCAAAACCCCGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db 601 ACCCAAAACCCCGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Oy 781 CCTACCAACCCCTGCGAGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Db 661 CCTACCAACCCCTGCGAGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Oy 841 GTAGCTCTCTGACCTCTCTCAATCAAGAGACTCTCCAGTGTCTAATGAGGAGCTCT 900
Db 721 GTAGCTCTCTGACCTCTCTCAATCAAGAGACTCTCCAGTGTCTAATGAGGAGCTCT 780
Oy 901 TTTCTTTCTGTCGTCCTTTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 960
Db 781 TTTCTTTCTGTCGTCCTTTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 840
Oy 961 AGACCCGACTACTACCAAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 841 AGACCCGACTACTACCAAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Oy 1021 AAAACAAGGGGGTCTTCTGAGGCTCTCCATATTAAGTTCAGGAGATCTGAGTGTGTA 1080
Db 961 AAAACAAGGGGGTCTTCTGAGGCTCTCCATATTAAGTTCAGGAGATCTGAGTGTGTA 960
Oy 1081 CAATTGACTCTGAGCTCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 961 CAATTGACTCTGAGCTCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Oy 1141 AATCAATTAATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db 1021 AATCAATTAATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Oy 1201 AGTCATGTCGCAATTTCTTTCTCTGAGCAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGATC 1260
Db 1081 AGTCATGTCGCAATTTCTTTCTCTGAGCAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGATC 1140
Oy 1261 GCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1320
Db 1141 GCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1200
Oy 1321 GCTGTCTGTCAAGTCCCGGAAAGAACTACGGGAGCTGAGATCTTTCCAGCCCGGAGT 1380
Db 1201 GCTGTCTGTCAAGTCCCGGAAAGAACTACGGGAGCTGAGATCTTTCCAGCCCGGAGT 1260
Oy 1381 ACTTACCATCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1261 ACTTACCATCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Oy 1441 AGCAGTACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500

```

```

Db 1321 AGCAGTACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
Oy 1501 TCTTACCAAAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Db 1381 TCTTACCAAAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Oy 1561 TGAGCTGATGAG 1572
Db 1441 TGAGCTGATGAG 1452

RESULT 7
AAV48329
ID AAV48329 standard; cDNA, 1527 BP.
XX
XX AC
XX AAV48329;
XX
XX DT
XX 20-NOV-1998 (first entry)
XX
XX DE
XX MiniMUC1 gene.
XX
XX KM
XX ss; MUC1; recombinant pox virus; cytotoxic T-lymphocyte; immunogen;
XX tumour; tumour-associated antigen.
XX
XX OS
XX Homo sapiens.
XX
XX PH
XX Key
XX CDS 1..1527
XX FT /*tag= a
XX FT /product= "MiniMUC1 protein"
XX
XX PN
XX MO9837095-A2.
XX
XX PD
XX 27-AUG-1998.
XX
XX PF
XX 24-FEB-1998; 98MO-US03693.
XX
XX PR
XX 24-FEB-1997; 97US-0038253.
XX
XX PA
XX (DAND ) DNA PARBER CANCER INST INC.
XX (THER-) THERION BIOLOGICS CORP.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX PI
XX Grtitz L, Kantor J, Kufe D, Panicali D, Schlom J;
XX
XX DR
XX WPI; 1998-467492/40.
XX
XX DR
XX P-PSDB; AAW77233.
XX
XX PT
XX New recombinant pox virus for tumour therapy - comprises DNA
XX encoding an immunogenic mini-MUC1 fragment comprising 5-25 MUC1
XX tandem repeat units
XX
XX PS
XX Example 1; Page 21-22; 42pp; English.
XX
XX CC
XX The immunogenic mini-MUC1 fragment was created from MUC1 tandem repeat
XX units for inclusion in a recombinant pox virus (RPV). The RPV was used
XX in a pharmaceutical composition also containing an immunomodulator to
XX generate MUC1 specific cytotoxic T-lymphocytes. The recombinant pox
XX virus therefore encodes an immunogenic MUC1 fragment that does not
XX undergo significant genetic deletion, thereby providing an unexpectedly
XX stable and immunogenic pox virus. They can be used to prevent or treat
XX tumours expressing MUC1 tumour-associated antigens.
XX
XX SO
XX Sequence 1527 BP; 296 A; 573 C; 351 G; 307 T; 0 other.

Query Match 80.0%; Score 1257; DB 19; Length 1527;
Best Local Similarity 90.8%; Pred. No. 6e-263;
Matches 1992; Conservative 0; Mismatches 90; Indels 51; Gaps 3;

Oy 58 ATGACACCGGACACCAAGTCTCTTCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 111
Db 1 ATGACACCGGACACCAAGTCTCTTCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 60

```

```

Qy 112 -----ACAGTTGTTACAGTCTTGGTCTCAATGACAGGCTCTACCCCA 150
Db 61 ACCACAGCCCCCTAAACCCCGACAGTGTGTTACGGGTTCTGGTCTACAGAGCTCTACCCCA 120
Qy 151 GGTGGAGAAAGAGAGACTTGGGTACCCAGAGAAAGTTCAATGGCCAGCTCTACTGAGAG 210
Db 121 GGTGGAGAAAGAGAGACTTGGGTACCCAGAGAAAGTTCAATGGCCAGCTCTACTGAGAG 180
Qy 211 AATGTGTAGATGACAGAGCCGTAAGCTCTCAAGCCAGAGCCCGGTTCAAGCTCTCTC 270
Db 181 AATGTGTAGATGACAGAGCTTGATA-----TCGAATCCGGTGTCCGGGCTCCAGC 234
Qy 271 ACCACTGAGAGACAGATGTCACTCTGGCCCGGACAGGAAACAGCTTCAAGTTCAAGCT 330
Db 235 GCGCCCCAGCCAGCGGTGTCACTTGGCCCGGAGACACAGGCCGGGCTCCAGC 294
Qy 331 GGCACCTGGGAGACAGATGTCACTTGGTCCAGTCCAGAGCCAGCCCTGGGCTCCAGC 390
Db 295 GCGCCCCAGCCAGCGGTGTCACTTGGCCCGGAGACACAGGCCGGGCTCCAGC 354
Qy 391 ACCCGCAGCCAGACAGATGTCACTTGAAGCCCGGAGCAAGAGCCAGCTCTCAAC 450
Db 355 GCGCCCCAGCCAGCGGTGTCACTTGGCCCGGAGACACAGGCCGGGCTCTCAAC 414
Qy 451 GCGCCCCAGCCAGCGGTGTCACTTGGCCCGGAGCAAGAG----- 492
Db 415 GCAACCCAGCCAGCGGTGTCACTTGGCCCGGAGACACAGGCCGGGCTCTCAAC 474
Qy 493 CCGCCCGCCGGGCTCTCAACCGCCCGGCGCCAGAGTGTCACTTGGCCCGGAGCAAGAG 552
Db 475 CCGCCCGCCGGGCTCTCAACCGCCCGGCGCCAGCGGTGTCACTTGGCCCGGAGCAAGAG 534
Qy 553 CCGCCCGCCGGGCTCTCAACCGCCCGGCGCCAGAGTGTCACTTGGCCCGGAGCAAGAG 612
Db 535 CCGCCCGCCGGGCTCTCAACCGCCCGGCGCCAGCGGTGTCACTTGGCCCGGAGCAAGAG 594
Qy 613 CCGCCCTTGGGCTCTCAACCGCCCTCAAGTGTCAAGTGTCACTTGGGCTCTCAAGCTCTGCA 672
Db 595 CCGCCCTTGGGCTCTCAACCGCCCTCAAGTGTCAAGTGTCACTTGGGCTCTCAAGCTCTGCA 654
Qy 673 TCAGAGCTCAGCTTCTCACTGTGTGACAAACGGGCACTCTCCAGGGCTTACACAAACCCCA 732
Db 655 TCAGAGCTCAGCTTCTCACTGTGTGACAAACGGGCACTCTCCAGGGCTTACACAAACCCCA 714
Qy 733 GGCAGCAGAGACACTCCACCCAGAGTTCACAGCAGCACTCTGATCTTACACAGCTCTG 792
Db 715 GGCAGCAGAGACACTCCATTTCTCAATTCACAGCAGCACTCTGATCTTACACAGCTCTG 774
Qy 793 GGCAGCAGAGACACTCCAGAGTTCACAGCAGCACTCTGATCTTACACAGCTCTG 852
Db 775 GGCAGCAGAGACACTCCATTTCTCAATTCACAGCAGCACTCTGATCTTACACAGCTCTG 834
Qy 853 ACCCTCTCAATCAACAGCACTTCTCCAGAGTGTCTAAGGAGTCTCTTCTTTCTCTG 912
Db 835 ACCCTCTCAATCAACAGCACTTCTCCAGAGTGTCTAAGGAGTCTCTTCTTTCTCTG 894
Qy 913 TCGTTTTCACATTTCAAACTCCAGTTTAATTCCTGTGAGAAATCCAGACCCAGATAC 972
Db 895 TCGTTTTCACATTTCAAACTCCAGTTTCTCTCTCTGTGAGAAATCCAGACCCAGATAC 954
Qy 973 TACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTCAGATTTTAACAAGGGGGT 1032
Db 955 TACCAAGAGCTGACAGAGACATTTCTCAAAATGTTTTCAGATTTTAACAAGGGGGT 1014
Qy 1033 TTTCTGGGCTCTCCAAATATTAAGTTCAAGCCAGAGATCTGTGTGTACAAATGACCTG 1092
Db 1015 TTTCTGGGCTCTCCAAATATTAAGTTCAAGGACAGAGATCTGTGTGTACAAATGACCTG 1074
Qy 1093 GCGTTCCGAGAGGATCAATGATGTCCAGCAGCTGTGAGACACAGTTCAATCAGATTA 1152
Db 1075 GCGTTCCGAGAGGATCAATGATGTCCAGCAGCTGTGAGACACAGTTCAATCAGATTA 1134

```

```

Qy 1153 ACCGAGAGAGGCTCTGATATTAACCTGACAGATCTCAGAGCTCAGGCTGAGTCTGCGCA 1212
Db 1135 ACCGAGAGAGGCTCTGATATTAACCTGACAGATCTCAGAGCTCAGGCTGAGTCTGCGCA 1194
Qy 1213 TTTCTCTTCTCTGCCCCAGCTGTGGGGCTGGGGTGCAGAGCTGGGGGCAATGCGCTCTGGTG 1272
Db 1195 TTTCTCTTCTCTGCCCCAGCTGTGGGGCTGGGGTGCAGAGCTGGGGGCAATGCGCTCTGGTG 1254
Qy 1273 CTGGTCTGTGTTCTGGTTGGCTGAGCAATGTCTATCTCATTTGCTTGGCTGTCTGTAG 1332
Db 1255 CTGGTCTGTGTTCTGGTTGGCTGAGCAATGTCTATCTCATTTGCTTGGCTGTCTGTAG 1314
Qy 1333 TGCAGCCAGAAAGAACTACAGGGGAGCTGAGACATCTTTCAGAGCCGGGATATCTTACATCT 1392
Db 1315 TGCAGCCAGAAAGAACTACAGGGGAGCTGAGCAATCTTTCAGAGCCGGGATATCTTACATCT 1374
Qy 1393 ATGAGAGAGTACCCACCTACCAACCAATGGGCGCTAATGTGCCCCCTTACAGAGTACGAT 1452
Db 1375 ATGAGAGAGTACCCACCTACCAACCAATGGGCGCTAATGTGCCCCCTTACAGAGTACGAT 1434
Qy 1453 CGTAGCCCTTATGAAAGTTTCTGCAAGTATAGTGGACAGAGCTCTTTACACAAC 1512
Db 1435 CGTAGCCCTTATGAAAGTTTCTGCAAGTATAGTGGACAGAGCTCTTTACACAAC 1494
Qy 1513 CCAGCAGTGGAGGCACTTCTGCAACTTGTAG 1545
Db 1495 CCAGCAGTGGAGGCACTTCTGCAACTTGTAG 1527

RESULT 8
AABD0394
ID AABD0394 standard; DNA; 1737 BP.
XX
AC AABD0394;
XX
XX 29-AUG-2000 (first entry)
XX
Db ubiquitin-B, coli LacI-human Mucin 1 fusion protein encoding DNA #5.
XX
KW Ubiquitin; LacI; beta-galactosidase; fusion protein;
KW human; Mucin 1; MUC-1; tumour; pMR30 expression vector;
KW anti-tumour; therapy; immune response; cytosolic; vaccine; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..1737
FT /tag= a
FT /product= "ubiquitin-LacI-MUC-1 fusion protein"
FT /tag= b
FT /label= UBILacI_DNA
FT /note= "includes ubiquitin-B, coli LacI fusion DNA"
FT /tag= c
FT /note= "Human MUC-1 partial DNA that corresponds to
FT nucleotides 136-1497 of the EMBL sequence J05581 with
FT two stop codons"
FT
FT
XX
XX W0200025827-R2.
XX
XX 11-MAY-2000.
XX
XX 18-OCT-1999; 99MO-EP07874.
XX
XX 30-OCT-1998; 98IT-MI02330.
XX
XX (MENA ) MENARINI RICERCHE SPA.
XX
XX Parente D, Di Massimo AM, De Santis R;
XX
XX WPI; 2000-365410/31.
XX
XX

```


PN MO200025827-A2.
 XX
 PD 11-MAY-2000.
 XX
 PF 18-OCT-1999; 99MO-BP07874.
 XX
 PR 30-OCT-1998; 98IT-MT02330.
 XX
 PA (MENA) MENARINI RICERCHE SPA.
 XX
 PI Parente D, Di Massimo AM, De Santis R;
 XX
 DR MPI; 2000-365410/31.
 DR P-PSDB; AAV71024.
 XX
 PT Composition containing one or more DNA molecules encoding fragments of
 PT a Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in
 PT anti-tumor therapy -
 XX
 PS Claim 16; Fig 5; 56pp; English.
 XX
 CC The present sequence is a DNA encoding a fragment of human Mucin 1
 CC (MUC-1) antigenic protein which is overexpressed in tumor cells. The
 CC sequence was obtained by PCR from plasmids pMR5165, pMR5167, pMR5168
 CC and pMR5169 which contain MUC-1 DNA from B720 tumour cells. It
 CC corresponds to nucleotides 136-1497 of the EMBL sequence
 CC 105581 with a start codon and two stop codons. The present sequence
 CC is cloned into a pMR530 expression vector and used in pharmaceutical
 CC composition e.g. vaccine for inducing an antigen-specific anti-tumour
 CC immune response. Composition containing this DNA molecule is useful in
 CC anti-tumour therapy of patients affected with tumours characterised
 CC by high MUC-1 expression.
 XX
 SO Sequence 1371 BP; 300 A; 450 C; 321 G; 300 T; 0 other:
 Query Match 77.0%; Score 1210.2; DB 21; Length 1371;
 Best Local Similarity 92.1%; Pred. No. 8.2e-253;
 Matches 1312; Conservative 0; Mismatches 53; Indels 60; Gaps 1;
 QY 119 TTACAGGTTCTGATGATGAGCTTACCCAGGTGAGAAAGAGAGACTTGTGCTACCC 178
 DB 2 TGACAGGTTCTGATGATGAGCTTACCCAGGTGAGAAAGAGAGACTTGTGCTACCC 61
 QY 179 AGAGAGGTTCTGATGATGAGCTTACCCAGGTGAGAAAGAGAGACTTGTGCTACCC 218
 DB 62 AGAGAGGTTCTGATGATGAGCTTACCCAGGTGAGAAAGAGAGACTTGTGCTACCC 121
 QY 239 TCTCCAGGCTCAGGCTCCGCTTCAAGCTCTCCAGCTCAGGAGACAGATGCTACCTG 258
 DB 122 TCTCCAGGCTCAGGCTCCGCTTCAAGCTCTCCAGCTCAGGAGACAGATGCTACCTG 181
 QY 299 CCCCAGGCTCAGGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 358
 DB 182 CCCCAGGCTCAGGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 241
 QY 359 TCCCAAGTACAGGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 418
 DB 242 TCCCAAGTACAGGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 301
 QY 419 CCCCAGGACAAAGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 478
 DB 302 CCCCAGGACAAAGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 340
 QY 479 CCCCAGGACAAAGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 538
 DB 341 CCCCAGGACAAAGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 361
 QY 539 CCCCAGGACAAAGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 598
 DB 362 CTCGAGATATCAGGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 421
 QY 599 CCCCAGGACAAAGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 658

DB 422 CCCCAGGACAAAGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 481
 QY 659 CCCCAGGACAAAGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 718
 DB 482 CTCGAGATATCAGGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 541
 QY 719 CCCCAGGACAAAGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 778
 DB 542 CCCCAGGACAAAGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 601
 QY 779 CTCCTACCAACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 838
 DB 602 CTCCTACCAACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 661
 QY 839 CCGTACCTCTCTCAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 898
 DB 662 CCGTACCTCTCTCAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 721
 QY 899 CTTCTCTTCT 958
 DB 722 CTTCTCTTCT 781
 QY 959 CAGACACGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 1018
 DB 782 CAGACACGACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 841
 QY 1019 ATAAACAGGAGGCTTCT 1078
 DB 842 ATAAACAGGAGGCTTCT 901
 QY 1079 TACAATGACTCTGAGCTTCT 1138
 DB 902 TACAATGACTCTGAGCTTCT 961
 QY 1139 TCAATGACTCTGAGCTTCT 1198
 DB 962 TCAATGACTCTGAGCTTCT 1021
 QY 1199 TGAATGACTCTGAGCTTCT 1258
 DB 1022 TGAATGACTCTGAGCTTCT 1081
 QY 1259 TCGAGCTGAGTCTGAGCTTCT 1318
 DB 1082 TCGAGCTGAGTCTGAGCTTCT 1141
 QY 1319 TGGCTGCTGAGTCTGAGCTTCT 1378
 DB 1142 TGGCTGCTGAGTCTGAGCTTCT 1201
 QY 1379 ATACCTACCT 1438
 DB 1202 ATACCTACCT 1261
 QY 1439 CTAGAGTACCT 1498
 DB 1262 CTAGAGTACCT 1321
 QY 1499 TCTCTTACCAAAACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 1543
 DB 1322 TCTCTTACCAAAACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 1366

RESULT 10
 AEN95623
 ID AEN95623 standard; DNA; 4139 BP.
 XX
 AC AEN95623;
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE Gene #2121 used to diagnose liver cancer.
 XX

KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KM metastatic liver tumour; cytostatic; expression profile; disease state;
 KM disease progression; drug toxicity; drug efficacy; drug metabolism.
 XX Homo sapiens.
 OS
 PN NC00229103-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 02-OCT-2001; 2001MO-US0589.
 XX
 PR 02-OCT-2000; 2000US-237054P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 DR WPI; 2002-426119/45.
 XX
 PT Diagnosing and detecting the progression of liver cancer,
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,
 PT involves detecting the level of expression of two or more genes in a
 PT liver tissue sample
 XX
 BS Claim 1; SEQ ID NO 2121; 298bp; English.
 XX
 CC The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytostatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WPI
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SO Sequence 4139 BP; 632 A; 1910 C; 1055 G; 542 T; 0 other;
 Query Match 75.7%; Score 1190.2; DB 24; Length 4139;
 Best Local Similarity 94.8%; Pred. No. 2.2e-248;
 Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
 QY 249 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGCCCGGCGAC
 DB 2545 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGCCCGGCGAC
 QY 309 GGAACCACTTCAAGTTCAGTCTGACCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 2605 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 369 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 2665 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 429 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 2725 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 489 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 2785 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 549 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 2845 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA

QY 609 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 2905 CAGCCCGGCTTCAAGCTCTTCACTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 669 TGCATCAGGCTTCAAGCTCTTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 2965 TGCATCAGGCTTCAAGCTCTTCACTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 729 CCGAGCAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3025 CCGAGCAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 789 CCGAGCAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3085 CCGAGCAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 849 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3145 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 909 CCGAGCAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3205 CCGAGCAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 969 CCGAGCAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3265 CCGAGCAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 1029 CCGAGCAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3325 CCGAGCAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 1089 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3385 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 1149 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3445 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 1209 GCGATTTCTTCTGAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3505 GCGATTTCTTCTGAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 1269 GCGATTTCTTCTGAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3565 GCGATTTCTTCTGAGGAGAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 1329 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3625 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 1389 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3685 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 1449 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3745 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 QY 1509 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 DB 3805 TGTCACTCTTCAATCAGATGTCATCTGAGGAGAGATGTCATCTGAGTCCAGTCA
 RESULT 11
 ID ABLE7071 standard; DNA; 4139 BP.
 XX
 AC ABLE7071;
 XX
 DT 15-MAY-2002 (first entry)
 XX

DE Thyroid cancer related gene sequence SEQ ID NO:5408.
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KM stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KM cytostatic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;
XX gene; ds.
OS Homo sapiens.
PN WO200194629-A2.
PD 13-DEC-2001.
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-23133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 02-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PT Soppe DR, Weaver Z;
XX
XX WPI: 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 5408; 44bp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical

CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL1664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer.
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms tumour.
XX
XX Sequence 4139 BP; 632 A; 1910 C; 1055 G; 542 T; 0 other;
SO
Query Match 75.7%; Score 1190.2; DB 24; Length 4139;
Best Local Similarity 94.8%; Pred. No. 2.2e-248;
Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 249 CAGCCCGGTTCAAGCTCTCCACCACTCAAGGACAGAGATGTCACCTGCCCCGAC
Db 2545 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
QY 309 GGAACCAAGCTTCAGGTTCAAGTCCACCTGGGGACAGAGATGTCACCTGCCCCGAC
Db 2605 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
QY 369 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
Db 2665 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
QY 429 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
Db 2725 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
QY 489 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
Db 2785 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
QY 549 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
Db 2845 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
QY 609 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
Db 2905 CAGGCGCGCCCGGGCTCCACCGCCCCCAGCCACGGTGTACTCGGCCCGGACAC
QY 669 TGGATCAGGCTCAAGCTTCTAAGTGTGACAGGACGACCTGCGAGGGCTACACAC
Db 2965 TGGATCAGGCTCAAGCTTCTAAGTGTGACAGGACGACCTGCGAGGGCTACACAC
QY 729 CCAGGCGAGGAGGACCTCAGCCAGGATTCAGGAGGACGACCTGAGTACCTTACAC
Db 3025 CCAGGCGAGGAGGACCTCAGCCAGGATTCAGGAGGACGACCTGAGTACCTTACAC
QY 789 CTTGCGAGGATGACGACCAAGATGATGACGATGACGATGACGATGACGATGAC
Db 3085 CTTGCGAGGATGACGACCAAGATGATGACGATGACGATGACGATGACGATGAC
QY 849 TCTCAGCTCTCATACAGACACTTCCCAAGTGTCTACTGAGGTCTTTCTTTT
Db 3145 TCTCAGCTCTCATACAGACACTTCCCAAGTGTCTACTGAGGTCTTTCTTTT
QY 909 CCGTCTTTTACATTTCAACCTCAGATTAATCTCTCGGAGATTCACAGACCA
Db 3205 CCGTCTTTTACATTTCAACCTCAGATTAATCTCTCGGAGATTCACAGACCA
QY 969 CTAACACAGAGAGTGAAGAGAGATTTGAAAGTTTGGAGATTTAAAGAG 1028
Db 3265 CTAACACAGAGAGTGAAGAGAGATTTGAAAGTTTGGAGATTTAAAGAG 3324

```

OY 1029 GGGTTTCTGGAGCTCTCCATATTAAGTTCAAGGCCAGAGTCTGGTGGTACAAATTGAC 1088
DB 3325 GGGTTTCTGGAGCTCTCTCCATATTAAGTTCAAGGCCAGAGTCTGGTGGTACAAATTGAC 3384
OY 1089 TCTGGCTCTCCGAGAGGATCATATGTCACGACGTGAGACACAGTTCAATCAGTA 1148
DB 3385 TCTGGCTCTCCGAGAGGATCATATGTCACGACGTGAGACACAGTTCAATCAGTA 3444
OY 1149 TAAACGGAAGCAGCTCTCGATATTAACCTGACGATCTGACGTCAGCTGATCATGT 1208
DB 3445 TAAACGGAAGCAGCTCTCGATATTAACCTGACGATCTGACGTCAGCTGATCATGT 3504
OY 1209 GCCATTTCTTTCTCTGCGCACTCTGGGGCTGGGGTCCGAGGTCGAGGCTGCT 1268
DB 3505 GCCATTTCTTTCTCTGCGCACTCTGGGGCTGGGGTCCGAGGTCGAGGCTGCT 3564
OY 1269 GGTGCTGGTCTGTGTTCTGTTGCGTGGCCATTTGCTATCTCATTTGCGCTGTCTG 1328
DB 3565 GGTGCTGGTCTGTGTTCTGTTGCGTGGCCATTTGCTATCTCATTTGCGCTGTCTG 3624
OY 1329 TGAAGCCGCGGAAAGATCAGGGCAGCTGGAATCTTCCAGCCCGGGAATACCTACCA 1388
DB 3625 TGAAGCCGCGGAAAGATCAGGGCAGCTGGAATCTTCCAGCCCGGGAATACCTACCA 3684
OY 1389 TCTATGAGGAGTACCCCACTACACACCACTGAGCGCTATGTGCCCTTAGCAGTAC 1448
DB 3685 TCTATGAGGAGTACCCCACTACACACCACTGAGCGCTATGTGCCCTTAGCAGTAC 3744
OY 1449 CGATCGTAGCCCTCATGAGAGGTTTCTGACGATGATGTTGAGCAGAGCTCTCTTACAC 1508
DB 3745 CGATCGTAGCCCTCATGAGAGGTTTCTGACGATGATGTTGAGCAGAGCTCTCTTACAC 3804
OY 1509 AAACCCAGCAGTGGCAGCACTTGGCCAATCTGTGAGG 1547
DB 3805 AAACCCAGCAGTGGCAGCGCTTGTCCAACTGTGAGG 3843
DB
RESULT 12
ABL67544
ID ABL67544 standard; DNA; 4139 BP.
XX
AC ABL67544;
XX
DT 15-MAY-2002 (first entry)
XX
XX
XX Thyroid cancer related gene sequence SEQ ID NO:5881.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
XX gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209511P.
PR 18-SEP-2000; 2000US-231133P.
PR 18-SEP-2000; 2000US-231617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234567P.
PR 22-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR

```

```

PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236103P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237315P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;
XX Soppet DR, Weaver Z;
XX WPI: 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 5881; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (II)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL6164
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic
XX activity and can be used in gene therapy. M1 can be used for screening
XX an anti-neoplastic agent, and can be used for producing a product which
XX is the data collected with respect to the anti-neoplastic agent as a
XX result of M1, and the data is sufficient to convey the chemical
XX structure and/or properties of the agent. M1 can be used in the
XX treatment of cancer such as colon, breast, stomach, lung, thyroid,
XX oesophageal, ovarian, kidney, prostate or pancreatic cancer,
XX adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
XX infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
XX carcinoma, papillary carcinoma and Wilms' tumour.
XX
XX Sequence 4139 BP; 632 A; 1910 C; 1055 G; 542 T; 0 other:
XX
XX Query March 75.7%; Score 1190.2; DB 24; Length 4139;
XX Best Local Similarity 94.8%; Pred. No. 2.2e-248; Indels 0;
XX Matches 1231; Conservative 0; Mismatches 66; Gaps 0;
OY 249 CAGCCCGGTTCAAGCTCTCTCCACATCAGGACAGATGTCATCTGCGCCGCGAC 308
DB 2545 CAGCGCGGCCCGGCTCCACCGCCGCCAGGCCAGGATGTCATCTGCGCCGCGAC 2604
OY 309 GGAACCACTTCAGTTCACTGTCACCTGAGGAGCAGAGATGTCACCTGCTCCAGTAC 368

```


PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.

XX (AVALU-) AVALON PHARM.

XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX WPI: 2002-188264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX Claim 1; SEQ ID 7221; 44pp; English.

XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in AB161664
CC to AB170110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Kilm's tumour.

XX Sequence 4139 BP; 632 A; 1910 C; 1055 G; 542 T; 0 other;

Query Match 75.7%; Score 1190.2; DB 24; Length 4139;
Best Local Similarity 94.8%; Pred. No. 2.2e-248;
Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 249 CAGCCCGGTTAGGCTCTTCAACATGAGAGAGATGCACTGCGCCCGGAC 308
DB 2545 CAGCGCGGCGGCTCCACCGCGCCCGAGCGGATGCACTGCGCCCGGAC 2604
QY 309 GGAACGAGTTGAGGTTGAGCTGCACTGCGGAGAGATGTCACCTGATCCAGTCA 368
DB 2605 CAGCGCGGCGGCTCCACCGCGCCCGAGCGGATGTCACCTGCGCCCGGAC 2664
QY 369 CAGCGCGGCTGCGGCTCCACCGCGCGGAGCGGATGTCACCTGAGCCCGGAC 428
DB 2665 CAGCGCGGCGGCTCCACCGCGCGGAGCGGATGTCACCTGCGCCCGGAC 2724
QY 429 CAAGCGAGCGCGGCTCCACCGCGCGGAGCGGATGTCACCTGCGCCCGGAC 488
DB 2725 CAGCGCGGCGGCTCCACCGCGCGGAGCGGATGTCACCTGCGCCCGGAC 2784
QY 489 CAGCGCGGCGGCTCCACCGCGCGGAGCGGATGTCACCTGCGCCCGGAC 548
DB 2785 CAGCGCGGCGGCTCCACCGCGCGGAGCGGATGTCACCTGCGCCCGGAC 2844
QY 549 CAGCGCGGCGGCTCCACCGCGCGGAGCGGATGTCACCTGCGCCCGGAC 608
DB 2845 CAGCGCGGCGGCTCCACCGCGCGGAGCGGATGTCACCTGCGCCCGGAC 2904
QY 609 CAGCGCGGCGGCTCCACCGCGCGGAGCGGATGTCACCTGCGCCCGGAC 668
DB 2905 CAGCGCGGCGGCTCCACCGCGCGGAGCGGATGTCACCTGCGCCCGGAC 2964
QY 669 TGCATCAGGCTTGAAGTTTGAAGTGAAGCAAGGCACTTGGCCAGGGCTACCAAC 728
DB 2965 TGCATCAGGCTTGAAGTTTGAAGTGAAGCAAGGCACTTGGCCAGGGCTACCAAC 3024

QY 729 CCGAGCGAGAAAGACATCCACCGAGCATTCACGACCACTTGTATCTCTACAC 788
DB 3025 CCGAGCGAGAAAGACATCCACCGAGCATTCACGACCACTTGTATCTCTACAC 3084
QY 789 CTTGCGAGCGCTGACACCAAGATGATGCGATGACATGACATGACAGGATCTCC 848
DB 3085 CTTGCGAGCGCTGACACCAAGATGATGCGATGACATGACATGACAGGATCTCC 3144
QY 849 TCTCAGCTCTCCATACACAGCATCTTCCCGAGTTGTCTACGTGGGGTCTCTTTT 908
DB 3145 TCTCAGCTCTCCATACACAGCATCTTCCCGAGTTGTCTACGTGGGGTCTCTTTT 3204
QY 909 CTTGCTTTTTCATTTTAAACCTCCAGTTTAAATCTCTCGAAGATCCGACCCCA 968
DB 3205 CTTGCTTTTTCATTTTAAACCTCCAGTTTAAATCTCTCGAAGATCCGACCCCA 3264
QY 969 CTACTACAGAGCTGACAGAGACATTTCTGAATGTTTTTTCAGATTTTAAACAG 1028
DB 3265 CTACTACAGAGCTGACAGAGACATTTCTGAATGTTTTTTCAGATTTTAAACAG 3324
QY 1029 GGGTTTTTTCGCGCTCTCCATATTTAGTTGAGCGAGATCTGATGATGATGAC 1088
DB 3325 GGGTTTTTTCGCGCTCTCCATATTTAGTTGAGCGAGATCTGATGATGATGAC 3384
QY 1089 TCTGGCTTCCGAGAGGATACATCAATGTCACAGCTGAGACACAGTTCAATCGTA 1148
DB 3385 TCTGGCTTCCGAGAGGATACATCAATGTCACAGCTGAGACACAGTTCAATCGTA 3444
QY 1149 TAAAGGAGAGAGCGCTCCGATTTAACTGAGATCTCAGAGCTCAGGATGATGT 1208
DB 3445 TAAAGGAGAGAGCGCTCCGATTTAACTGAGATCTCAGAGCTCAGGATGATGT 3504
QY 1209 GCGATTTCTTTTCTGCGGAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1268
DB 3505 GCGATTTCTTTTCTGCGGAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 3564
QY 1269 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1328
DB 3565 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3624
QY 1329 TCAATGCGCGCGAAGAACTACGAGGAGCTGAGATCTTCCAGCCCGGATACCA 1388
DB 3625 TCAATGCGCGCGAAGAACTACGAGGAGCTGAGATCTTCCAGCCCGGATACCA 3684
QY 1389 TCTATAGAGAGATACCCCACTACCAACCATGAGGCTATGTCCTTACAGATAC 1448
DB 3685 TCTATAGAGAGATACCCCACTACCAACCATGAGGCTATGTCCTTACAGATAC 3744
QY 1449 CGATGTAAGCCCTTATGAGAGGTTTCTGCAAGTATGTTGAGAGGCTTCTTAC 1508
DB 3745 CGATGTAAGCCCTTATGAGAGGTTTCTGCAAGTATGTTGAGAGGCTTCTTAC 3804
QY 1509 AAACCGAGAGTGGAGGCACTTTCGCAACTGTAGGG 1547
DB 3805 AAACCGAGAGTGGAGGCACTTTCGCAACTGTAGGG 3843

RESULT 14
ABK09797
ID ABK09797 standard; cDNA; 4139 BP.
XX
XX ABK09797;
DT 14-MAR-2002 (first entry)
XX
XX Human ovarian tumour protein encoding cDNA #330.
XX Human; ovarian tumour protein; cancer; cytostatic; immunostimulant; ss;
XX gene therapy; CD4+ T cell; CD8+ T cell; PCR primer.
XX Homo sapiens.
XX
XX MO200190154-A2.
PN

Thu May 8 18:53:06 2003

us-09-658-621b-1.rng

Page 20

Db 1021 GTGGCAGCCACTTCTGCACTTGTAGGGGACGTGCGCCCTCTGAGCTGAGTGG 1074

Search completed: May 8, 2003, 15:04:24
Job time : 447 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: May 8, 2003, 14:54:17 ; Search time 2408 Seconds
(without alignments)
10572.801 Million cell updates/sec

Title: US-09-658-621B-1

Perfect score: 1572
Sequence: 1 gaattccctgcgtcgttga.....tcgcccctgcagctcgtcgtg 1572

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em.estba:*
2: em.esthm:*
3: em.estlm:*
4: em.estma:*
5: em.estov:*
6: em.estpl:*
7: em.estro:*
8: em.estc:*
9: gb.estc1:*
10: gb.estc2:*
11: gb.estc3:*
12: gb.estc4:*
13: gb.estc5:*
14: gb.estc6:*
15: em.estfun:*
16: em.estom:*
17: gb.ges:*
18: em.ges.hum:*
19: em.ges.juv:*
20: em.ges.pln:*
21: em.ges.vit:*
22: em.ges.fun:*
23: em.ges.mam:*
24: em.ges.mus:*
25: em.ges.other:*
26: em.ges.pro:*
27: em.ges.tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARYS

Result No.	Score	Query Match	Length	ID	Description
1	701.8	44.6	874	BQ918845	BQ918845 AGENCOURT
2	701.8	44.6	944	BQ928921	BQ928921 AGENCOURT
3	659.8	42.0	682	AL543598	AL543598 AL543598
4	653.8	41.6	670	BM617771	BM617771 K-EST0084
5	642.4	40.9	798	BQ740690	BQ740690 602631536
6	635.8	40.4	664	BM794696	BM794696 K-EST0076

7	621	39.5	670	12	BF338440
8	611.4	38.9	621	14	BM629853
9	594.4	37.8	604	14	BM650782
10	589.8	37.5	593	14	BM741056
11	573.4	36.5	595	14	BM795113
12	557.2	35.4	559	10	AM387919
13	555.8	35.4	559	14	AM793904
14	554.4	35.3	981	12	BM774910
15	543.8	34.6	629	12	BM791452
16	541	34.3	950	12	BM775831
17	539.8	34.3	927	12	BM698098
18	530.8	33.8	604	14	BM791359
19	527.4	33.5	529	14	BM633236
20	525.4	33.4	524	14	BM631319
21	520.8	31.9	524	14	BM622982
22	494.6	31.5	566	10	AM369405
23	487.8	31.0	580	9	AI245417
24	486.8	31.0	502	14	BM746832
25	478.4	30.4	480	14	BM652127
26	474	30.2	722	9	AL543556
27	466.2	29.7	471	14	BM745435
28	459.8	29.2	759	14	BM682837
29	454.6	28.9	770	14	BM691344
30	447.8	28.5	725	13	BI761034
31	447.6	28.5	472	14	BM759495
32	447.6	28.5	485	13	BI014455
33	444.6	28.3	760	14	BM693267
34	428.4	27.3	437	14	BM686872
35	425.4	27.1	438	14	BM620231
36	424.4	26.9	424	14	BM797973
37	419.2	26.7	590	9	AA429320
38	419.2	26.7	719	14	BM684687
39	414	26.3	442	10	AM579047
40	408.2	26.0	745	10	AM001137
41	407.8	25.9	711	14	BM674802
42	405.8	25.8	707	14	BM446679
43	403.4	25.7	908	9	AA826345
44	398.4	25.3	630	9	AI905920
45	398.2	25.3	485	10	AM369421

ALIGNMENTS

RESULT 1
LOCUS BQ918845 874 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT 877143 NIH_MGC_40 Homo sapiens CDNA clone IMAGE:6384674
5', mRNA Sequence.
ACCESSION BQ918845
VERSION BQ918845.1 GI:22333543
KEYWORDS EST.

SOURCE

ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS (Bases 1 to 874)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCPD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LINC2580 row: b column: 03
High quality sequence stop: 621.
Location/Qualifiers
1. 874

FEATURES

source

```

/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:634674"
/clone.lib="NIH_MGC_40"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: prostate; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

```

BASE COUNT 175 a 252 c 220 g 226 t 1 others

Query Match 44.6%; Score 701.8; DB 14; Length 874;
 Best Local Similarity 99.6%; Pred. No. 8.4e-147;
 Matches 703; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 867 CAGCACTTCTCCCAAGTGTCTACTGGGGCTCTTTTCTTTTCTGCTTTTCATTTTC 926
Db 3 CAGCACTTCTCCCAAGTGTCTACTGGGGCTCTTTTCTTTTCTGCTTTTCATTTTC 62
Qy 927 AAACCTCAGTTTAAATCTCTCTGGAAGATCCGACACCGACTACTACCAAGAGCTGCA 986
Db 63 AAACCTCAGTTTAAATCTCTCTGGAAGATCCGACACCGACTACTACCAAGAGCTGCA 122
Qy 987 GAGAGACATTTCTGAATGTTTTTTCAGATTATTAACAAGAGGGGTTTTTCGGGCTCTC 1046
Db 123 GAGAGACATTTCTGAATGTTTTTTCAGATTATTAACAAGAGGGGTTTTTCGGGCTCTC 182
Qy 1047 CAATTTAAGTTCAGGCGAGGATCTGTGGTGTACATTTGACTCTGGGCTTCGAGAAAG 1106
Db 183 CAATTTAAGTTCAGGCGAGGATCTGTGGTGTACATTTGACTCTGGGCTTCGAGAAAG 242
Qy 1107 TACCATCAATGTCCAGACGTGGAGACACAGTTCAATCACTATTAACAGAAAGAGCTTC 1166
Db 243 TACCATCAATGTCCAGACGTGGAGACACAGTTCAATCACTATTAACAGAAAGAGCTTC 302
Qy 1167 TCGATATAACCTGACGATCTCAGACGTCAGGCGTGAAGTCAATTCCTTTCTCTGC 1226
Db 303 TCGATATAACCTGACGATCTCAGACGTCAGGCGTGAAGTCAATTCCTTTCTCTGC 362
Qy 1227 CCAAGTCTGGGGCTGGGGTGCAGAGGCTGGGGGATCGAGCTCTGGTGTCTGTCTTC 1286
Db 363 CCAAGTCTGGGGCTGGGGTGCAGAGGCTGGGGGATCGAGCTCTGGTGTCTGTCTTC 422
Qy 1287 GGTGGGCTGGGCAATGTCTACTGCTGTGGTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1346
Db 423 GGTGGGCTGGGCAATGTCTACTGCTGTGGTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 482
Qy 1347 CTACGGGAGCTGGAGCATCTTTCAGACCCGGGATACCTACATCTTATGACGATACCC 1406
Db 483 CTACGGGAGCTGGAGCATCTTTCAGACCCGGGATACCTACATCTTATGACGATACCC 542
Qy 1407 CACCTACACACCCCATGGGGGCTATGTGCCCCCTAGAGTACGATCTTACGCTTATGA 1466
Db 543 CACCTACACACCCCATGGGGGCTATGTGCCCCCTAGAGTACGATCTTACGCTTATGA 602
Qy 1467 GAAGGTTCTCAGAGTAAATGTGGAGAGGCTCTCTTACCAAAACCCAGAGAGGAGAG 1526
Db 603 GAAGGTTCTCAGAGTAAATGTGGAGAGGCTCTCTTACCAAAACCCAGAGAGGAGAG 662
Qy 1527 CACTTCTGCAACTTGTGAGGGGAGCGTGGCCCTTGTAGCTGAGTGG 1572
Db 663 CACTTCTGCAACTTGTGAGGGGAGCGTGGCCCTTGTAGCTGAGTGG 708

```

RESULT 2
 BQ928921 944 bp mRNA linear EST 20-AUG-2002
 LOCUS BQ928921

DEFINITION AGENCOURT 10032855 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6481446
 5' mRNA Sequence.
 ACCESSION BQ928921
 VERSION BQ928921.1 GI:22343952
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: DCM/DTP
 cDNA Library Preparation: Rubin laboratory
 DNA Sequencing by: Agencourt BioScience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLMB at:
<http://image.llnl.gov>
 Plate: LUCW2662 row: n column: 07
 High quality sequence stop: 687.
 Location/Qualifiers

FEATURES
 source 1..944
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6481446"
 /clone.lib="NIH_MGC_40"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /notes="Organ: prostate; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 182 a 271 c 252 g 237 t 2 others
 ORIGIN
 Query Match 44.6%; Score 701.8; DB 14; Length 944;
 Best Local Similarity 99.6%; Pred. No. 8.4e-147;
 Matches 703; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 867 CAGCACTTCTCCCAAGTGTCTACTGGGGCTCTTTTCTTTTCTGCTTTTCATTTTC 926
Db 3 CAGCACTTCTCCCAAGTGTCTACTGGGGCTCTTTTCTTTTCTGCTTTTCATTTTC 62
Qy 927 AAACCTCAGTTTAAATCTCTCTGGAAGATCCGACACCGACTACTACCAAGAGCTGCA 986
Db 63 AAACCTCAGTTTAAATCTCTCTGGAAGATCCGACACCGACTACTACCAAGAGCTGCA 122
Qy 987 GAGAGACATTTCTGAATGTTTTTTCAGATTATTAACAAGAGGGGTTTTTCGGGCTCTC 1046
Db 123 GAGAGACATTTCTGAATGTTTTTTCAGATTATTAACAAGAGGGGTTTTTCGGGCTCTC 182
Qy 1047 CAATTTAAGTTCAGGCGAGGATCTGTGGTGTACATTTGACTCTGGGCTTCGAGAAAG 1106
Db 183 CAATTTAAGTTCAGGCGAGGATCTGTGGTGTACATTTGACTCTGGGCTTCGAGAAAG 242
Qy 1107 TACCATCAATGTCCAGACGTGGAGACACAGTTCAATCACTATTAACAGAAAGAGCTTC 1166
Db 243 TACCATCAATGTCCAGACGTGGAGACACAGTTCAATCACTATTAACAGAAAGAGCTTC 302
Qy 1167 TCGATATAACCTGACGATCTCAGACGTCAGGCGTGAAGTCAATTCCTTTCTCTGC 1226
Db 303 TCGATATAACCTGACGATCTCAGACGTCAGGCGTGAAGTCAATTCCTTTCTCTGC 362
Qy 1227 CCAAGTCTGGGGCTGGGGTGCAGAGGCTGGGGGATCGAGCTCTGGTGTCTGTCTTC 1286

```

Thu May 8 18:53:10 2003

us-09-658-621b-1.rst

Page 3

[illegible]

RESULT 3	AL543598	682 bp	mrna	linear	EST 16-FEB-2001
LOCUS	AL543598				
DEFINITION	AL543598.1 IT1.NFL006.Pl2.Homo.sapiens cDNA clone CS001006Y07.5				
ACCESSION	AL543598				
VERSION	AL543598.1	GI:12876077			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheta; Primates; Catarrhini; Homiidae; Homo.				
REFERENCE	1 (bases 1 to 682)				
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polyes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Genoscope				

genoscope - centre national de séquençage
BP 191 91006 EVRY cedex - France
Email: secretaire@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..682
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSOD1006Y07"
/clone_1b="LTI NFL006 PL2"
/tissue_type="Placenta"
/notes="Vector: PCWVSORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCWVSORT 6 vector. library was normalized. library was constructed by a life Technologies Contact : Peng liang life Technologies, a division of invitrogen 9600 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : liliang@lifestech.com URL : http://fulllength.invitrogen.com"

```

	Query Match	42.0%	Score 659.8	DB %	Length 682
	Best Local Similarity	98.8%	Pred. No. 2.1e-137		
	Matches 675	Conservative	0	Mismatches 7	Indels 1
Qy	595	TCGCGCCCGGAAACAAGGCGCGGCGCTTGGGCTGCACCGGCGCTTCAGTGCACAAATGTCACC	654		
Db	1	TCGCGCCCGGAAACAAGGCGCGGCGCTTGGGCTGCACCGGCGCTTCAGTGCACAAATGTCACC	60		
Qy	655	TCGCGCTCAGGCTGTGATCAGGCTCAGCTTAACTCTGTGTGCAACAAGCAACTCTTGCC	714		

Dp	61	TCGGCTCAGAGCTCGCATCGAGCTCAGCTTTACTGTGGTGCACAAACGGCACCTCTGCC	120
Qy	725	AGGGCTTACCAACATCCCGACGACGAAGAGCATTCACACCATTCCTCCAGCCACCACTCT	774
Dp	121	AGGGCTTACCAACATCCCGACGACGAAGAGCATTCATTTCTCAATTTCCAGCCACCACTCT	180
Qy	775	GATATCTCTACCACTCCCTTGTGCAGCGATTTGACACCAAGATGTGCAGTATGACATCTACAT	834
Dp	181	GATATCTCTACCACTCCCTTGTGCAGCGATTTGACACCAAGATGTGCAGTATGACATCTACAT	240
Qy	835	AGACAGGTAACCTCTCTCACTCTCTCCATTCACATCAACGACCTTCTCCCAAGTTGTCTAC	894
Dp	241	AGACAGGTAACCTCTCTCACTCTCTCCATTCACATCAACGACCTTCTCCCAAGTTGTCTAC	300
Qy	895	GTCCTTTCTTTTCTCTGTCTTTTACATTTTCAAACTTCAGTTAATTCCTCTGTGAA	954
Dp	301	GTCCTTTCTTTTCTCTGTCTTTTACATTTTCAAACTTCAGTTAATTCCTCTGTGAA	360
Qy	955	GATCCAGACCCGACTACTACCAAGATGTGCAGAGACATTTGTGAAATGTTTGTGAG	1014
Dp	361	GATCCAGACCCGACTACTACTACCAAGATGTGCAGAGACATTTGTGAAATGTTTGTGAG	420
Qy	1015	ATTATTAATAAAGGGGGTTTTCTGGGCTCTCCCAATTAATGTTAGGCGAGAGATCTGTG	1074
Dp	421	ATTATTAATAAAGGGGGTTTTCTGGGCTCTCCCAATTAATTAAGTTTCAAGCGAGAGATCTGTG	480
Qy	1075	GTGTGTCAATTGACTCTGGGCTTTCGAGAAAGTACCATCATATGTCCACGACGTGGAGCA	1134
Dp	481	GTGTGTCAATTGACTCTGGGCTTTCGAGAAAGTACCATCATATGTCCACGACGTGGAGCA	539
Qy	1135	CAGTTCAATCAGATTAATAAAGAAACGACACCTCTCGATTAATACGACGATCTCAGACCTC	1194
Dp	540	CAGTTCAATCAGATTAATAAAGAAACGACACCTCTCGATTAATACGACGATCTCAGACCTC	599
Qy	1195	AGCGTAGTCAATGAGCAATTCCTTTCTGTGCCATCTGGGGGTCTGGGGGTGCAGGCTGG	1254
Dp	600	AGCGTAGTCAATGAGCAATTCCTTTCTGTGCCATCTGGGGGTCTGGGGGTGCAGGCTGG	659
Qy	1255	GGCATCGCGCTGTGTGTCTGTGT	1277
Dp	660	GGCATCGCGCTGTGTGTCTGTGTGT	682

LOCUS	670 bp	mRNA	linear	EST 06-MAR-2002
DEFINITION	X:EB000771.4	S21SNUS2061	Homo sapiens	CDNA clone S21SNUS2061-2-D07
ACCESSION	KJ850044243			
VERSION	5			
KEYWORDS	"5', mRNA sequence."			
SOURCE	BM837771			
ORGANISM	BM817771.1	GI:19174184		
	EST.			
	human.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 670)			
AUTHORS	Kim,N.S., Hahn,Y.Y., Oh,J.Y., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,			
	Oh,K.U., Choeng,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and			
	Kim,Y.S.			
TITLE	21C Frontier Korean EST Project 2001			
JOURNAL	Unpublished (2002)			
COMMENT	Contact: Kim YS			
	Genome Research Center			
	Korea Research Institute of Bioscience & Biotechnology			
	55 Boseun-dong Yusong-gu, Daejeon 305-333, South Korea			
	Tel: +82-42-860-4470			
	Fax: +82-42-860-4409			
	Email: yongsung@mail.kribb.re.kr			
	Plate: 12 row: D column: 07			
FEATURES	High quality sequence stop: 670.			
source	Location/Qualifiers			
	1..670			
	/organism="Homo sapiens"			

```

/db_xref="taxon:9606"
/clone="S21SN052081-2-D07"
/clone_lib="S21SN052081"
/sex="F"
/tissue_type="Stomach"
/cell_line="SNU-520"
/lab_host="Top10F"
/notes="Organ: Stomach; Vector: pTZ19RP1; Site 1: EcoRI; Site 2: NotI; The poly (A) RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then deacapped with tabacco acid pyrophosphatase (TAP). The deacapped intrac mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoR I which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library. After analyzing and sequencing about 2,000 ~ 3,000 colonies in original cDNA library, the abundant cDNAs were selected and amplified by PCR reaction using vector region primer including T7 promoter as 5' primer and Nid14 as 3' primer. The PCR products were used as template for synthesis of biotinylated single stranded RNA by in vitro transcription reaction. The synthesized RNA probes were hybridized with antisense single stranded cDNAs prepared from original library and incubated with avidin-gel. After removing DNA-RNA hybrids by centrifuge, the subtracted cDNA libraries were constructed by transformation of the remaining DNA into competent cells E. coli Top10F with electroporation method."
BASE COUNT      146 a      188 c      170 g      166 t
ORIGIN
Query Match      41.6%; Score 653.8; DB 14; Length 670;
Best Local Similarity 99.7%; Pred. No. 4.5e-136;
Matches 655; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 916 TTTCACATTGAACTCCAGTTTAACTCTCTCGAGAAAGATCCGACCGACTACTAC 975
Db 1 TTTCACATTGAACTCCAGTTTAACTCTCTCGAGAAAGATCCGACCGACTACTAC 60
Oy 976 CAAGAGCTCCAGAGACATTTCTGAAATGTTTTCGACATTTTAAACAAGGGGTTT 1035
Db 61 CAAGAGCTCCAGAGACATTTCTGAAATGTTTTCGACATTTTAAACAAGGGGTTT 120
Oy 1036 CTGGGCTCTCCATATTAAGTTCAAGCCAGAGATCTGTGTGACATTAATGCTG 1095
Db 121 CTGGGCTCTCCATATTAAGTTCAAGCCAGAGATCTGTGTGACATTAATGCTG 180
Oy 1096 TTCCGAGAAAGTACATCATATGTCACAGCGTGAAGACACAGTTCAATGATTAACG 1155
Db 181 TTCCGAGAAAGTACATCATATGTCACAGCGTGAAGACACAGTTCAATGATTAACG 240
Oy 1156 GAAGAGCCTCTCCATTAACCTGACGATCTCAGCGCGTGCAGTGCATGTCGCAATT 1215
Db 241 GAAGAGCCTCTCCATTAACCTGACGATCTCAGCGCGTGCAGTGCATGTCGCAATT 300
Oy 1216 CTTTCTTCCGCGAGTCTGGGGCTGGGGTGCAGCGTGGGGCATGCGCTGTGGTGG 1275
Db 301 CTTTCTTCCGCGAGTCTGGGGCTGGGGTGCAGCGTGGGGCATGCGCTGTGGTGG 360
Oy 1276 GTCTGTGTTCTGGTTGGCTGGCCATTTGCTATCTCATGCTTGGCTGTCTGCAATG 1335
Db 361 GTCTGTGTTCTGGTTGGCTGGCCATTTGCTATCTCATGCTTGGCTGTCTGCAATG 420
Oy 1336 CGCGAAAGAACTAGCGAGCTGACATCTTCCAGCCGGGATCACTAATCACTATG 1395

```

```

Db 421 CGCGAAAGAACTAGCGAGCTGACATCTTCCAGCCGGGATCACTAATCACTATG 480
Oy 1396 AGCGATACCCCACTTCAACACCCATGCGCTATGTCGCCCTGACGATACCATCT 1455
Db 481 AGCGATACCCCACTTCAACACCCATGCGCTATGTCGCCCTGACGATACCATCT 540
Oy 1456 AGCCCTATGAGAAAGTTTCTCAGATTAATGTGTGGAGCAGGCTCTCTTACACAAACCA 1515
Db 541 AGCCCTATGAGAAAGTTTCTCAGATTAATGTGTGGAGCAGGCTCTCTTACACAAACCA 600
Oy 1516 GCAGTGGAGCAGCACTTCTGCAAACTTTAGAGGAGCAGTGCCTCTGTAGCTGAGTGG 1572
Db 601 GCAGTGGAGCAGCACTTCTGCAAACTTTAGAGGAGCAGTGCCTCTGTAGCTGAGTGG 657

RESULT 5
Bg740690      798 bp      mRNA      linear      EST 15-MAY-2001
LOCUS      602631536F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4776632 5',
DEFINITION      mRNA sequence.
ACCESSION      Bg740690
VERSION      Bg740690.1 GI:14051343
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 798)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LHAM10629 row: h column: 09
High quality sequence stop: 798.
Location/Qualifiers
source      1..798
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4776632"
/clone_lib="NCI CGAP_Skn3"
/lab_host="DH10B (T1_phase-resistant)"
/notes="Organ: Skin; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: oligo dt. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      177 a      224 c      198 g      199 t
ORIGIN
Query Match      40.9%; Score 642.4; DB 12; Length 798;
Best Local Similarity 99.7%; Pred. No. 1.7e-133;
Matches 654; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Oy 817 GCCAGTAGACTCACTAGCAGGATGACCTCTCAGCTCCATCAATCAAGCACTTCT 876
Db 1 GCCAGTAGACTCACTAGCAGGATGACCTCTCTCACTCTCCATCAAGCACTTCT 60
Oy 877 CCCAGTTCTACTAGGGGCTCTTTCTTCTGCTTTCATCAATTGAAACCTCCG 936
Db 61 CCCAGTTCTACTAGGGGCTCTTTCTTCTGCTTTCATCAATTGAAACCTCCG 120
Oy 937 TTTAATCTCTCTGGAAGATCCAGACCG-ACCTACTACAAAGCTGCGAGAGCAT 995
Db 121 TTTAATCTCTCTGGAAGATCCAGACCGCACTAATCAAGAGCTGCGAGAGCAT 180

```

```

Qy 996 TTCTGAAATGTTTTTGAGATTTATTAACAAAGGGGTTTTCTGGGCTCTCCATATTAA 1055
Db 181 TTCTGAAATGTTTTTGAGATTTATTAACAAAGGGGTTTTCTGGGCTCTCCATATTAA 240
Qy 1056 GTTCAGAGCAGGATGTGGGTGATCAATTTGCTGGGCTTCGGAAGAGTACATCAA 1115
Db 241 GTTCAGAGCAGGATGTGGGTGATCAATTTGCTGGGCTTCGGAAGAGTACATCAA 300
Qy 1116 TTTCCACGACGTGGAGACACAGTTCAATCAATTAACAAAGAGGCTCTCCATATTAA 1175
Db 301 TGTCCACGACGTGGAGACACAGTTCAATCAATTAACAAAGAGGCTCTCCATATTAA 360
Qy 1176 CTTGAAGATCTCAGACGTGAGGTGATGTCATGTCATTTCTTCTGCTGAGTCTGG 1235
Db 361 CTTGAAGATCTCAGACGTGAGGTGATGTCATGTCATTTCTTCTGCTGAGTCTGG 420
Qy 1236 GAGCTGGGGTCCAGAGCTGGGGCATCGGGCTGGTGGTCTGTTCTGAGTCTGGGCT 1295
Db 421 GAGCTGGGGTCCAGAGCTGGGGCATCGGGCTGGTGGTCTGTTCTGAGTCTGGGCT 480
Qy 1296 GGGCATTTGTTCTCATTTGCTTGGCTGTCTGTCTCATGTCGGCGGAAGAACTACGGGCA 1355
Db 481 GGGCATTTGTTCTCATTTGCTTGGCTGTCTGTCTCATGTCGGCGGAAGAACTACGGGCA 540
Qy 1356 GCTGACATCTTTCCAGCCCGGAGTACTTACCATCTTATGAGCGAGTACCCCACTTACCA 1415
Db 541 GCTGACATCTTTCCAGCCCGGAGTACTTACCATCTTATGAGCGAGTACCCCACTTACCA 600
Qy 1416 CACCCATGGGCGCTATGTGCCCCCTAGCAGTACCGCATGTAGCCCCCTATGAGAGG 1471
Db 601 CACCCATGGGCGCTATGTGCCCCCTAGCAGTACCGCATGTAGCCCCCTATGAGAGG 656

```

RESULT 6

```

LOCUS BM794696 664 bp mRNA linear EST 05-MAR-2002
DEFINITION X-EST0076148 S22SNUI6n1 Homo sapiens cDNA clone S22SNUI6n1-48-A12
5', mRNA sequence.
ACCESSION BM794696
VERSION BM794696.1 GI:19142928
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 664)
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.U., Cheong, J.B., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 48 row: A column: 12
High quality sequence strip: 664.
Location/Qualifiers
1.664
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S22SNUI6n1-48-A12"
/clone_id="S22SNUI6n1"
/sex="F"
/tissue_type="Asclites"
/cell_type="Lymphoblast-like"
/lab_host="DH10B"
/note="Organ: Stomach; Vector: pUT73-Pac; Site: 1: ECoRI;
Site_2: NotI; The S22SNUI6 library was contributed by the

```

Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cells of SNU-16 culture. SNU-16 cell was obtained from Korean Cell Line Bank (KCLB). SNU-16 was established from arctic fluids of Korean patients by Park J.G. et al. (1990). Cancer Res 50: 2773-2780."

BASE COUNT 145 a 185 c 170 g 164 t

Query Match 40.4%; Score 635.8; DB 14; Length 664;

Best Local Similarity 99.5%; Pred. No. 4.9e-132;

Matches 648; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 922 ATTTCAACTCCAGTTTATTTCTCTCTGGAAGATCCGAGCAGCTACTACAAAG 981

Db 1 ATTTCAACTCCAGTTTATTTCTCTCTGGAAGATCCGAGCAGCTACTACAAAG 60

Qy 982 CTGAGAGAGACATTTCTGGAATGTTTTGACATTTATTAACAAAGGGGTTTTCTGGGC 1041

Db 61 CTGAGAGAGACATTTCTGGAATGTTTTG-AGATTTATTAACAAAGGGGTTTTCTGGGC 119

Qy 1042 CTCTCAATTTAAAGTTCAAGCCAGATCTGTGTGTATTAATTAATCTGCTTCCGA 1101

Db 120 CTCTCAATTTAAAGTTCAAGCCAGATCTGTGTGTATTAATTAATCTGCTTCCGA 179

Qy 1102 GAAAGTACATCAATGTCACAGACGTGAGACACAGTTCAATCAATTAACGGAAGA 1161

Db 180 GAAAGTACATCAATGTCACAGACGTGAGACACAGTTCAATCAATTAACGGAAGA 239

Qy 1162 GCTCTGATATTAATCTGACATCTCAGACGTGAGTCAATGTCATTTCTTTCC 1221

Db 240 GCTCTGATATTAATCTGACATCTCAGACGTGAGTCAATGTCATTTCTTTCC 299

Qy 1222 TCTGCCAGTCTGGGGCTGGGGTGGCGAGCTGGGGCATCGCGCTGGTGGTGGTCTGT 1281

Db 300 TCTGCCAGTCTGGGGCTGGGGTGGCGAGCTGGGGCATCGCGCTGGTGGTGGTCTGT 359

Qy 1282 GTTCGTGTGCGTGGCGATTTGTCTATCTCAATGCTTGGCTGTCTGAGTGC CGCGCA 1341

Db 360 GTTCGTGTGCGTGGCGATTTGTCTATCTCAATGCTTGGCTGTCTGAGTGC CGCGCA 419

Qy 1342 AAGAACTACGAGGAGCTGACATCTTTCAGCCCGGAGTACTTACCATCTTATGAGCGAG 1401

Db 420 AAGAACTACGAGGAGCTGACATCTTTCAGCCCGGAGTACTTACCATCTTATGAGCGAG 479

Qy 1402 TACCCCACTACACACACCCAGGAGCTATGTCGCCCTAGAGTACGATCTGAGGCC 1461

Db 480 TACCCCACTACACACACCCAGGAGCTATGTCGCCCTAGAGTACGATCTGAGGCC 539

Qy 1462 TATGAGAGGTTTCTGAGTATGTTGGAGCAGCTTCTTACCAAAACCGAGAGTG 1521

Db 540 TATGAGAGGTTTCTGAGTATGTTGGAGCAGCTTCTTACCAAAACCGAGAGTG 599

Qy 1522 GCAGCACTTCTGCAACTTGTAGGGGAGCGTCCCTGTGAGCTGAGTGG 1572

Db 600 GCAGCACTTCTGCAACTTGTAGGGGAGCGTCCCTGTGAGCTGAGTGG 650

RESULT 7
LOCUS BF338440 670 bp mRNA linear EST 22-NOV-2000
DEFINITION 602034094F2 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4162286
5', mRNA sequence.
ACCESSION BF338440
VERSION BF338440.1 GI:11284843
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 670)
NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINT at:
http://image.llnl.gov
Plate: LLM9496 row: k column: 23
High quality sequence stop: 669.

FEATURES
source
1. 670
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4182286"
/clone_1b="NCI CGAP Bm64"
/tissue_type="glioblastoma with EGFR amplification"
/lab_note="DH10B (TI phage-resistant)"
/notes="Organ: Brain; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.57 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP library."

BASE COUNT 151 a 221 c 136 g 162 t
ORIGIN

Query Match 39.5%; Score 621; DB 12; Length 670;
Best Local Similarity 98.3%; Pred. No. 1e-128;
Matches 638; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 598 GCGCCGAGCAACAGGCGCGCTTGGGCTCCAGCGCCCTCCAGTCCCAATGTCACCTGG 657
Db 22 GCGCCGAGCAACAGGCGCGCTTGGGCTCCAGCGCCCTCCAGTCCCAATGTCACCTGG 61
QY 658 GCGTCAAGCTCTGATAGAGCTGAGTCTTCTCTGATGCAACAGGCACTCTGCGCAG 717
Db 82 GCGTCAAGCTCTGATAGAGCTGAGTCTTCTCTGATGCAACAGGCACTCTGCGCAG 141
QY 718 GCTACCAACACCCAGCAGCAAGAGCACTCCACAGCAATCCAGCCACTCTGAT 777
Db 142 GCTACCAACACCCAGCAGCAAGAGCACTCCATTCATTCACAGCAGCACTCTGAT 201
QY 778 ACTGCTACCACTCTGCGAGCAGTACAGCAAGAGTATGCGAGTACATCAGCATAGC 837
Db 202 ACTGCTACCACTCTGCGAGCAGTACAGCAAGAGTATGCGAGTACATCAGCATAGC 261
QY 838 AAGGTAAGCTCTGCACTCCCTCCCAATGCAAGCACTTCCCGCAGTGTCTACTGAGGATC 897
Db 262 AAGGTAAGCTCTGCACTCCCTCCCAATGCAAGCACTTCCCGCAGTGTCTACTGAGGATC 321
QY 898 TCTTCTCTTCTCTGCTCTTCTTCAATTCACATTCGATTAATTCCTCTCTGGAAGAT 957
Db 322 TCTTCTCTTCTCTGCTCTTCTTCAATTCGATTAATTCCTCTCTGGAAGAT 381
QY 958 CCCAGACCGGACTCTACCAAGAGCTGAGAGAGCAATTCGAAATGTTTTTCAGATT 1017
Db 382 CCCAGACCGGACTCTACCAAGAGCTGAGAGAGCAATTCGAAATGTTTTTCAGATT 441
QY 1018 TTTAAACAAGGGGTTTCTGAGGCTCTCCAAATTTAAAGTTCAGGCGAGGATCTGTGGT 1077
Db 442 TTTAAACAAGGGGTTTCTGAGGCTCTCCAAATTTAAAGTTCAGGCGAGGATCTGTGGT 501
QY 1078 GTTAAATGATCTGCGGCTTCGAGAGGATCAATGATCCAGAGCTGAGAGCAAG 1137
Db 502 GTTAAATGATCTGCGGCTTCGAGAGGATCAATGATCCAGAGCTGAGAGCAAG 561
QY 1138 TTCAATCAAGTAAACGAGAGGCTCTGATTAACCTGACATCTCAGAGCTGAGC 1197
Db 562 TTCAATCAAGTAAACGAGAGGCTCTGATTAACCTGACATCTCAGAGCTGAGC 620
QY 1198 GTGAGTCAATGTCAGTCTCTCTCTGCGCAATCTGAGGCTGAGGATGC 1246

Db 621 GTGAGTCAATGTCAGTCTCTCTCTGCGCAATCTGAGGCTGAGGATGC 669

RESULT 8
LOCUS BM629853 621 bp mRNA linear EST 06-MAR-2002
DEFINITION K-EST0103028 S21SNUS2081 Homo sapiens cDNA clone S21SNUS2081-16-E03
5', mRNA sequence.
ACCESSION BM629853
VERSION BM629853.1 GI:19186262
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 621)
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoan-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 16 row: E column: 03
High quality sequence stop: 621.
Location/Qualifiers

FEATURES
source
1. 621
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S21SNUS2081-16-E03"
/clone_1b="S21SNUS2081"
/sex="F"
/tissue_type="Stomach"
/cell_type="floating aggregates"
/cell_line="SNU-520"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pTZ19BP1; Site_1: EcoRI;
Site_2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then deapped
with tobacco acid pyrophosphatase (TAP). The deapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dT-selected mRNA by
priming with dT-tailed vector. The dT-tailed vector was
adjusted to have about 60nt. The cDNA vector was
clonitized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 ~ 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and NdeI 14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNA prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transformation of the
remaining DNA into competent cells E. coli Top10F with
electroporation method."

BASE COUNT 137 a 188 c 134 g 162 t
ORIGIN

	OY	1216	CGTTCTGCGCCAGTCTGGGGGCTGGGGGACCAAGCTGGGGGCATGCCGTGTCGTCTG	1275
Dn	541	CGTTCTGCGCCAGTCTGGGGGCTGGGGGACCAAGCTGGGGGCATGCCGTGTCGTCTG		600
OY	1276			
Dn	601	GTCCT	604	
RESULT	10			
LOCUS	BM7A1056			
DEFINITION	K-EST001J425 S6SNU620 Homo sapiens cDNA clone S6SNU620-2-G04 5', mRNA sequence.	593 bp	mRNA	linear EST 01-MAR-2002
ACCESSION	BM7A1056			
VERSION	BM7A1056			
KEYWORDS	EST.			
SOURCE	BM7A1056.1 GI:19062385			
ORGANISM	human.			
AUTHORS	Homo sapiens Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Euthalia; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (bases 1 to 593)			
COMMENT	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.U., Cheong,J.E., Sohn,H.Y., Kim,D.M., Park,H.S., Kim,S. and Kim,Y.S.			
TITLE	21C Frontier Korean EST Project 2001			
JOURNAL	Unpublished (2002)			
COMMENT	Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Boseon-dong Yuseong-gu, Daejeon 305-333, South Korea Tel.: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongsung@mail.kribb.re.kr Place: 2 row: G column: 04 High quality sequence stop: 593.			

BASE COUNT	127 a	161 c	159 g	146 t
ORIGIN				

Query Match	37.5%	Score 589.8;	DB 14;	Length 593;
Best Local Similarity	99.7%;	Pred. No. 9.6e-122;		
Matches 591; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

977 AAGAGCTGACAGAGACATTTCGAAATGTTTTCAGATTATATAACAAGGGGTTTTTC 1036

Db	1	AAGAGCTGCAAGAGACATTTCTGAAATGTTTTGACGATTTATTTAAACAAGGAGGAGTTTC	60
Qy	1037	TGGAGCTCTCCAAATTAATTAAGTTCCAGGACAGGATCTGGTGGTACAATTGACCTCTGGACCT	1096
Db	61	TGGGCTCTCCAAATTAATTAAGTTCCAGGACAGGATCTGGTGGTACAATTGACCTCTGGACCT	120
Qy	1097	TCCGAAAGGATCCATCAATATTCACGACGCGAGACACAGTTTCAATAGATATTAACG	1156
Db	121	TCCGAAAGGATCCATCAATATTCACGACGCGAGACACAGTTTCAATAGATATTAACG	180
Qy	1157	AGCAGCCTCTCGATTAATCTTGACATCTTCAACGTCAAGCGTGAATCAGTCCATTTTC	1216
Db	181	AGCAGCCTCTCGATTAATCTTGACATCTTCAACGTCAAGCGTGAATCAGTCCATTTTC	240
Qy	1217	CTTTCTCTGCACATCTGGGAGCTGGGAGTCCAGGCTGGGAGCTCCGCTGGTGTCTG	1276
Db	241	CTTTCTCTGCACATCTGGGAGCTGGGAGTCCAGGCTGGGAGCTCCGCTGGTGTCTG	300
Qy	1277	TCGAGTCTGAGTTGGGCTGGGACATTTGTCACTCAATATGCTCTGGGCTGTCTGCAATGGC	1336
Db	301	TCGAGTCTGAGTTGGGCTGGGACATTTGTCACTCAATATGCTCTGGGCTGTCTGCAATGGC	360
Qy	1337	GCCGAAAGATCAACGGGACAGCTGACATCTTTCAGCCCGGATATCTACCATCTATGA	1396
Db	361	GCCGAAAGATCAACGGGACAGCTGACATCTTTCAGCCCGGATATCTACCATCTATGA	420
Qy	1397	GCGAGTACCCCACTACCAACCCATAGGAGCTATGTGCCCTTACGAGTACCGATCTGA	1456
Db	421	GCGAGTACCCCACTACCAACCCATAGGAGCTATGTGCCCTTACGAGTACCGATCTGA	480
Qy	1457	GCCCTATAGAAAGATTTCTGAGGATTAATGTTGGACAGACCTCTTAACCAAAACCGAG	1516
Db	481	GCCCTATAGAAAGATTTCTGAGGATTAATGTTGGACAGACCTCTCTTAACCAAAACCGAG	540
Qy	1517	CAGTGGACGACCTTCTGCAACCTTTATGAGGACACGTCCCTCTGACCTGAG	1569
Db	541	CAGTGGACGACCTTCTGCAACCTTTATGAGGACACGTCCCTCTGACCTGAG	593

Source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S21SNU520-27-E03"
/clone_lib="S21SNU520"
/sex="F"

```

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="S21SNU520-27-E03"  
/clone_lib="S21SNU520"  
/sex="F"
```


/tissue_type="Stomach"
/cell_type="Floating aggregates"
/cell_line="SNU-520"
/lab_host="Top10f"
/note="Organ: Stomach; Vector: pTZ19p1; Site 1: EcoRI;
Site 2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP) and the decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transfection of
competent cells E. coli Top10f by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT 137 a 183 c 124 g 151 t
ORIGIN

Query Match 36.5%; Score 573.4; DB 14; Length 595;
Best Local Similarity 98.8%; Pred. No. 4.5e-118;
Matches 588; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

676 GGCTCAGCTTCTACTCTGTGTGACACAGCACTCTGCAGGCTACCAACCCAGCC 735
1 GGCTCAGCTTCTACTCTGTGTGACACAGCACTCTGCAGGCTACCAACCCAGCC 60
726 AGCAAGAGCACTCCACGAGTATCCAGCAGCACTCTGATCTCTTACCCCTTGGC 795
61 AGCAAGAGCACTCCACGAGTATCCAGCAGCACTCTGATCTCTTACCCCTTGGC 120
796 AGCAAGAGCACTCCACGAGTATCCAGCAGCACTCTGATCTCTTACCCCTTGGC 855
121 AGCAAGAGCACTCCACGAGTATCCAGCAGCACTCTGATCTCTTACCCCTTGGC 180
856 TCTCCAGTACAGCACTCTTCCAGTATCTCTGAGGCTCTCTTCTTTTCTGTCT 915
181 TCTCCAGTACAGCACTCTTCCAGTATCTCTGAGGCTCTCTTCTTTTCTGTCT 240
916 TTTTCACTTTCAAACTCCAGTTTATTTCTCTCTGAGAAATCCAGCACTGACTAC 975
241 TTTTCACTTTCAAACTCCAGTTTATTTCTCTCTGAGAAATCCAGCACTGACTAC 300
976 CAAGAGCTGAGAGAGCACTTTGAAATGTTTGGAGATTATTAACAAGGGGTTT 1035
301 CAAGAGCTGAGAGAGCACTTTGAAATGTTTGGAGATTATTAACAAGGGGTTT 360
1036 CTGGGCTCTCCAAATTTAAGTTCAGGCGAGATCTGTGTGTGATCAATGACTGGCC 1095
361 CTGGGCTCTCCAAATTTAAGTTCAGGCGAGATCTGTGTGTGATCAATGACTGGCC 420
1096 TTTCCAGAAAGTATCATCATATGTCACAGAGTGAACAAGTTCATCATGATTAAG 1155
421 TTTCCAGAAAGTATCATCATATGTCACAGAGTGAACAAGTTCATCATGATTAAG 480
1156 GAAAGAGCTCTCTGATATTAACCTGAGAGTCTGAGAGTGAAGTCAATGAGCATTT 1215
481 GAAAGAGCTCTCTGATATTAACCTGAGAGTCTGAGAGTGAAGTCAATGAGCATTT 540
1216 CTTTCTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGT 1269
541 CTTTCTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGT 595

RESULT 12
AM387919/c 563 bp mRNA linear EST 04-FEB-2000
LOCUS AM387919
DEFINITION MR4-ST0119-071099-010-D02 ST0119 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM387919

VERSION AM387919.1 GI:6892578
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 563)
AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.
TITLE The PAPSP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPSP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR4&e2=MR4-ST0119-071099-010-D02&t3=1999-10-07&t4=1>)
Seq primer: puc 18 forward
High quality sequence, stop: 523.

FEATURES
source
1..563
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_11b="ST0119"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 140 a 151 c 157 g 114 t 1 others
ORIGIN

Query Match 35.4%; Score 557.2; DB 10; Length 563;
Best Local Similarity 99.3%; Pred. No. 1.9e-114;
Matches 559; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1004 TGTTTTTCAGATTATTAACAAGGGGTTTCTGGGCTCTCCAAATTAAGTTGAGGC 1063
563 TGTTTTTCAGATTATTAACAAGGGGTTTCTGGGCTCTCCAAATTAAGTTGAGGC 504
1064 CAGATCTGTGTGTGATCAATGACTCTGAGAGTATCCATCAATGTCAGG 1123
503 CAGATCTGTGTGTGATCAATGACTCTGAGAGTATCCATCAATGTCAGG 444
1124 AGTGTGAGACAGATTCAATGATTAACAAGAGACCTCTGATATTAACCTGACGA 1183
443 AGTGTGAGACAGATTCAATGATTAACAAGAGACCTCTGATATTAACCTGACGA 384
1184 TCTCAGAGCTGAGAGTCAATGATTAACAAGAGACCTCTGATATTAACCTGACGA 1243
383 TCTCAGAGCTGAGAGTCAATGATTAACAAGAGACCTCTGATATTAACCTGACGA 324
1244 TGCAGAGCTGAGAGTCAATGATTAACAAGAGACCTCTGATATTAACCTGACGA 1303
323 TGCAGAGCTGAGAGTCAATGATTAACAAGAGACCTCTGATATTAACCTGACGA 264
1304 TCTATCTATATGCTGTGTGTGATCAATGATTAACAAGAGACCTCTGATATTAACCTGACGA 1363
263 TCTATCTATATGCTGTGTGTGATCAATGATTAACAAGAGACCTCTGATATTAACCTGACGA 204
1364 TCTTTTCAGCCCGGAGTATCAATCTTATAGAGAGTATCCCACTTACCACTGATG 1423
203 TCTTTTCAGCCCGGAGTATCAATCTTATAGAGAGTATCCCACTTACCACTGATG 144
1424 GCGCTATATGTCCTCCCTTACAGTACCAATGATGATCCCTTATGAGAAAGTTTCTGACAGTA 1483

```

Db 143 GGGGCTATGTCCTCCCTTGGACGATCGTACCCCTTATGAGAGTTTCTGCAAGTA 84
Oy 1484 ATGGAGGACGACGAGCTCTCTTACACAAACCGACAGAGGACGACCTTCTGCACTTGG 1543
Db 83 ATGGAGGACGACGAGCTCTCTTACACAAACCGACAGAGGACGACCTTCTGCACTTGG 24
Oy 1544 AGGGGACGAGTGGCCCTCTGAGCT 1566
Db 23 AGGGGACGAGTGGCCCTCTGAGCT 1

RESULT 13
BM793904 559 bp mRNA linear EST 05-MAR-2002
LOCUS K-EST0074890 S22SNUI6n1 Homo sapiens cDNA clone S22SNUI6n1-32-A02
DEFINITION 5', mRNA sequence.
ACCESSION BM793904
VERSION BM793904.1 GI:19142136
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 559)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.U., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
TITLE Genome Research Center
JOURNAL Korea Research Institute of Bioscience & Biotechnology
COMMENT 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongjung@mail.kribb.re.kr
Plate: 32 row: A column: 02
High quality sequence stop: 559.
Location/Qualifiers
1..559
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S22SNUI6n1-32-A02"
/clone_1lb="S22SNUI6n1"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblastic-like"
/cell_line="SNUI-16"
/lab_host="DH10B"
/note="Organ: Stomach; Vector: pT73-pac; Site: 1; EcoRI;
Site_2: NotI; The S22SNUI6 library was contributed by the
Soares laboratory and it was constructed as described by
Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome
Research 6(9): 791-806. RNA was prepared from harvested
cells of SNUI-16 culture. SNUI-16 cell was obtained from
Korean Cell Line Bank (KCLB). SNUI-16 was established from
ascitic fluids of Korean patients by Park J.G. et al.
(1990), Cancer Res 50: 2773-2780."
BASE COUNT 120 a 154 c 145 g 140 t
ORIGIN
Query Match 35.4%; Score 555.8; DB 14; Length 559;
Best Local Similarity 99.6%; Pred. No. 3.9e-114;
Matches 557; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 1064 CAGGATCTGGTGGTGAATTAATTAAGTCTGGGCTTCCGAGAGGTAACATATGTCAGG 1123
Db 121 CAGGATCTGGTGGTGAATTAATTAAGTCTGGGCTTCCGAGAGGTAACATATGTCAGG 180
Oy 1124 AGGTGAGACAGGTTTCAATCGATGATTAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 1183
Db 181 AGGTGAGACAGGTTTCAATCGATGATTAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 240
Oy 1184 TCTCAGAGCTCAGGTTGAGTCAATGTCGATTTCTCTGTCGAGGCTGGGGCTGGGG 1243
Db 241 TCTCAGAGCTCAGGTTGAGTCAATGTCGATTTCTCTGTCGAGGCTGGGGCTGGGG 300
Oy 1244 TGCAGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1303
Db 301 TGCAGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 360
Oy 1304 TCTATCTGATGCTTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 1363
Db 361 TCTATCTGATGCTTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 420
Oy 1364 TCTTTCAGCCCGGATACCTACATCTCTATGACGAGTACCCCACTTACACACCATG 1423
Db 421 TCTTTCAGCCCGGATACCTACATCTCTATGACGAGTACCCCACTTACACACCATG 480
Oy 1424 GGGGCTATGTCCTCCCTTACGAGTACCGATCGTACCCCTTATGAGAGGTTTCTGCAAGTA 1483
Db 481 GGGGCTATGTCCTCCCTTACGAGTACCGATCGTACCCCTTATGAGAGGTTTCTGCAAGTA 540
Oy 1484 ATGGTGGACGAGGCTCTC 1502
Db 541 ATGGTGGACGAGGCTCTC 559

RESULT 14
LOCUS BG774910 981 bp mRNA linear EST 15-MAY-2001
DEFINITION 60264983P1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:4761054 5',
ACCESSION BG774910
VERSION BG774910.1 GI:14045227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 981)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: gcrabs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM612 row: c column: 07
High quality sequence stop: 874.
Location/Qualifiers
1..981
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4761054"
/clone_1lb="NIH MGC 40"
/tissue_type="Carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOTB1; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACAG(G). Library constructed by

```



```

Db      61  |||||
GACTGAGCTTCCGAGAGGTACCATCATGTCCACAGCTGGAGACAGTTCAATCA 120

QY      1146  GTATATAACGAGAGAGAGCTCTCGATATATACCTGACGATCTCAGACGTCAAGGTGATCA 1205
|||||
Db      121  GTATATAACGAGAGAGAGCTCTCGATATATACCTGACGATCTCAGACGTCAAGGTGATGA 180
|||||
QY      1206  TGTGCGCATTTCTTTCTCTGCGCAGTCTGGGGCTGGGGTGCAGGCTGGGGCATGCGCT 1265
|||||
Db      181  TGTGCGCATTTCTTTCTCTGCGCAGTCTGGGGCTGGGGTGCAGGCTGGGGCATGCGCT 240
|||||
QY      1266  GCTGGTCTGCTGTGTGTCTGTGCTGGGCTGTGCTGTCTATCTCATTTGCTTGGCTGT 1325
|||||
Db      241  GCTGGTCTGCTGTGTGTCTGTGCTGGGCTGTGCTGTCTATCTCATTTGCTTGGCTGT 300
|||||
QY      1326  CTGTCAGTGCGCGGAAAGAACTACGGGAGCTGGACATCTTCCAGCCGGGATACCTA 1385
|||||
Db      301  CTGTCAGTGCGCGGAAAGAACTACGGGAGCTGGACATCTTCCAGCCGGGATACCTA 360
|||||
QY      1386  CCATCCGATGAGCGAGTACCCCACTACGACACCCATGGGGGCTATGTGCCCTAGCAG 1445
|||||
Db      361  CCATCCGATGAGCGAGTACCCCACTACGACACCCATGGGGGCTATGTGCCCTAGCAG 420
|||||
QY      1446  TACCGATGAGCGCGCTATGAGAGGTTCTGAGGTAAAGTGGCAGCAGCTCTCTTA 1505
|||||
Db      421  TACCGATGAGCGCGCTATGAGAGGTTCTGAGGTAAAGTGGCAGCAGCTCTCTTA 480
|||||
QY      1506  CACAAACCCAGAGTGGCAGCCACTTCTGCAACTTGTAGGGGCACTGCGCCCTGAGC 1565
|||||
Db      481  CAGAAACCCAGAGTGGCAGCCACTTCTGCAACTTGTAGGGGCACTGCGCCCTGAGC 540
|||||
QY      1566  TTAGTGG 1572
|||||
Db      541  TTAGTGG 547

```

Search completed: May 8, 2003, 16:57:38
 Job time : 2462 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 15:26:48 ; Search time 766 Seconds
(without alignments)
9678.616 Million cell updates/sec

Title: US-09-658-621B-1
Perfect score: 1572
Sequence: 1 gaattccctgagctgctgaa.....tcgcccctgagctgagtg 1572

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 6120518 seqs, 235805216 residues

Total number of hits satisfying chosen parameters: 12241036

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents NA New:
1: /cgn2_6/pdata/2/pna/PCT_NEW_COMB.seq:
2: /cgn2_6/pdata/2/pna/US06_NEW_COMB.seq:
3: /cgn2_6/pdata/2/pna/US07_NEW_COMB.seq:
4: /cgn2_6/pdata/2/pna/US08_NEW_COMB.seq:
5: /cgn2_6/pdata/2/pna/US09_NEW_COMB.seq:
6: /cgn2_6/pdata/2/pna/US09_NEW_COMB.seq:
7: /cgn2_6/pdata/2/pna/US09_NEW_COMB.seq:
8: /cgn2_6/pdata/2/pna/US10_NEW_COMB.seq:
9: /cgn2_6/pdata/2/pna/US10_NEW_COMB.seq:
10: /cgn2_6/pdata/2/pna/US60_NEW_COMB.seq:
11: /cgn2_6/pdata/2/pna/US60_NEW_COMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1572	100.0	1572	5	US-09-658-621B-1
2	1481.2	94.2	1721	9	US-10-342-887-775
3	1412.2	89.8	1804	9	US-10-348-119-113
4	1412.2	89.8	1804	11	US-60-459-782-30
5	1412.2	89.8	1804	11	US-60-427-982-30
6	1412.2	89.8	1823	9	US-10-101-510-339
7	1411.8	89.8	1549	8	US-10-170-235-24915
8	1409.8	89.7	1549	11	US-60-452-680-471
9	1409.8	89.7	1549	11	US-60-453-135-343
10	1409.8	89.7	1549	11	US-60-453-050-343
11	1190.2	75.7	4139	1	PCT-US02-19668A-110
12	952.2	60.7	1353	8	US-10-170-235-24364
13	952.2	60.6	1353	11	US-60-452-680-474
14	952.2	60.6	1353	11	US-60-453-135-346
15	952.2	60.6	1353	11	US-60-453-050-346
16	848.6	54.0	1272	8	US-10-170-235-24365
17	847.8	53.9	1272	11	US-60-452-680-473
18	847.8	53.9	1272	11	US-60-453-135-345
19	847.8	53.9	1272	11	US-60-453-050-345
20	750	47.7	8186	11	US-60-422-176-21
21	632.8	40.3	1173	8	US-10-170-235-24366
22	631.2	40.2	1173	11	US-60-452-680-472

23	631.2	40.2	1173	11	US-60-453-135-344	Sequence 344, App
24	631.2	40.2	1173	11	US-60-453-050-344	Sequence 344, App
25	584	37.2	2166	9	US-10-101-510-326	Sequence 326, App
26	584	37.2	2235	9	US-10-144-771-7897	Sequence 7897, App
27	459	29.2	5660	9	US-10-311-455-2330	Sequence 2330, App
28	328.4	20.9	330	1	PCT-US02-34777-9	Sequence 9, App1
29	328.4	20.9	330	9	US-10-283-017-9	Sequence 9, App1
30	282.2	15.0	573	9	US-10-101-510-49	Sequence 49, App1
31	235.4	15.0	5660	9	US-10-311-455-2329	Sequence 2329, App
32	212.8	13.5	216	1	PCT-US02-34777-941	Sequence 941, App
33	212.8	13.5	216	9	US-10-283-017-941	Sequence 1210, App
34	212.8	13.5	216	9	US-10-283-017-941	Sequence 1210, App
35	212.8	13.5	216	9	US-60-452-680-29235	Sequence 29235, A
36	200.6	12.8	201	11	US-60-453-135-19343	Sequence 19343, A
37	200.6	12.8	201	11	US-60-453-050-19343	Sequence 19343, A
38	200.6	12.8	201	11	US-60-452-680-29237	Sequence 29237, A
39	199	12.7	201	11	US-60-452-680-29242	Sequence 29242, A
40	199	12.7	201	11	US-60-453-135-19345	Sequence 19345, A
41	199	12.7	201	11	US-60-453-135-19350	Sequence 19350, A
42	199	12.7	201	11	US-60-453-050-19345	Sequence 19345, A
43	199	12.7	201	11	US-60-453-050-19350	Sequence 19350, A
44	199	12.7	201	11	US-60-453-050-19350	Sequence 19350, A
45	180.4	11.5	201	11	US-60-452-680-29236	Sequence 29236, A

ALIGNMENTS

RESULT 1
US-09-658-621B-1
Sequence 1, Application US/09658621B
GENERAL INFORMATION:
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Heukamp, Lukas Carl
APPLICANT: Offringa, Rienk
APPLICANT: Melief, Cornelia Johanna Maria
APPLICANT: Acres, Bruce
APPLICANT: Thomas, Mireille
TITLE OF INVENTION: MGC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621B
CURRENT FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1572
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (58) .. (1542)
US-09-658-621B-1
Query Match 100.0%; Score 1572; DB 5; Length 1572;
Best Local Similarity 100.0%; Pred. No. 7.1e-296;
Matches 1572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCCTGCTGCTTGAATCTTTTGGCCCTCCACCACTTACACACACG 60
DB 1 GAATTCCTGCTGCTTGAATCTTTTGGCCCTCCACCACTTACACACACG 60
QY 61 ACACGGGACCCAGCTCTCTTCTTCTGCTGCTCTTCTTCTTCTTCTTCTT 120
DB 61 ACACGGGACCCAGCTCTCTTCTTCTGCTGCTCTTCTTCTTCTTCTTCTT 120
QY 121 ACAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
DB 121 ACAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180

```
Db 121 ACAGGTTCTGTCATGCAAGCTTACCCAGGTGAGAAAAGAGACATTGGGCTACCCAG 180
Oy 181 AGAAGTTCAGTGGCCAGGCTCTACTGAGAGAAATGCTGTGAGTATGACAGAGGCTACTTC 240
Db 181 AGAAGTTCAGTGGCCAGGCTCTACTGAGAGAAATGCTGTGAGTATGACAGAGGCTACTTC 240
Oy 241 TCCAGGCCACAGCCCGGTTTCAGGCTCTTCCACACTCAGAGGACAGAGTGTCACTCTGGCC 300
Db 241 TCCAGGCCACAGCCCGGTTTCAGGCTCTTCCACACTCAGAGGACAGAGTGTCACTCTGGCC 300
Oy 301 CCGGACGAGAAACAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 360
Db 301 CCGGACGAGAAACAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 360
Oy 361 CCAAGTCAACAGAGCCAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 420
Db 361 CCAAGTCAACAGAGCCAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 420
Oy 421 CCGGACGAGAAACAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 480
Db 421 CCGGACGAGAAACAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 480
Oy 481 CCGGACGAGAGCCAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 540
Db 481 CCGGACGAGAGCCAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 540
Oy 541 CCGGACGAGAGCCAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 600
Db 541 CCGGACGAGAGCCAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 600
Oy 601 CCGGACGAGAGCCAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 660
Db 601 CCGGACGAGAGCCAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 660
Oy 661 TCAAGGCTCTGCAATCAGGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 720
Db 661 TCAAGGCTCTGCAATCAGGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 720
Oy 721 ACCGACAAACCCAGCCAGGACAGGACCTCCACAGATTTCCAGCCAGCAGCTCTGATCT 780
Db 721 ACCGACAAACCCAGCCAGGACAGGACCTCCACAGATTTCCAGCCAGCAGCTCTGATCT 780
Oy 781 CCGGACGAGAGCCAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 840
Db 781 CCGGACGAGAGCCAGCTTTCAGTTCAGCTGCACCTGAGGAGACAGATGTCACTCTGGCTC 840
Oy 841 GTCACCTCTCTCACTCTCTCAATCAGAGACTTTCCTCCAGCTTTCCTCTGAGAGATCC 900
Db 841 GTCACCTCTCTCACTCTCTCAATCAGAGACTTTCCTCCAGCTTTCCTCTGAGAGATCC 900
Oy 901 TTTCTTTTCTGTCTTTTCAATTTCAAACTCCAGTTTAAATTCCTCTGAGAGATCC 960
Db 901 TTTCTTTTCTGTCTTTTCAATTTCAAACTCCAGTTTAAATTCCTCTGAGAGATCC 960
Oy 961 AGCAGCGACTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTCAGATTTAT 1020
Db 961 AGCAGCGACTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTCAGATTTAT 1020
Oy 1021 AAACAGAGGGGTTTTCGAGGCTTTCATATTAAGTTCAAGCCAGATCTGTGGTGA 1080
Db 1021 AAACAGAGGGGTTTTCGAGGCTTTCATATTAAGTTCAAGCCAGATCTGTGGTGA 1080
Oy 1081 CAATTGACTCTGAGCTTTCAGAGAGTACCAATGTCACGAGCTGAGAGACAGATTC 1140
Db 1081 CAATTGACTCTGAGCTTTCAGAGAGTACCAATGTCACGAGCTGAGAGACAGATTC 1140
Oy 1141 AATCAGATTAATAAGAGAGAGCTCTCGATTAATCTGAGAGATCTGAGAGCTGAGCGTG 1200
Db 1141 AATCAGATTAATAAGAGAGAGCTCTCGATTAATCTGAGAGATCTGAGAGCTGAGCGTG 1200
Oy 1201 AGTCAAGTGTCAATTTCTTTCTGAGCCAGCTGAGGAGCTGAGGAGCTGAGGAGCTG 1260
Db 1201 AGTCAAGTGTCAATTTCTTTCTGAGCCAGCTGAGGAGCTGAGGAGCTGAGGAGCTG 1260
```

```
Oy 1261 GCGGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db 1261 GCGGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Oy 1321 GCTGTCTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
Db 1321 GCTGTCTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
Oy 1381 ACTTACCATCTTATGAGGAGTACCCAGCTTACCAACACCAATGAGGAGCTTATGTGCCCCCT 1440
Db 1381 ACTTACCATCTTATGAGGAGTACCCAGCTTACCAACACCAATGAGGAGCTTATGTGCCCCCT 1440
Oy 1441 AGCAGTACCGATCTGAGGAGTACCCAGCTTATGAGAGATTTCTGAGGATTAATGAGGAGCTG 1500
Db 1441 AGCAGTACCGATCTGAGGAGTACCCAGCTTATGAGAGATTTCTGAGGATTAATGAGGAGCTG 1500
Oy 1501 TCTTACCAAAACCCAGGAGTACCCAGCTTATGAGAGATTTCTGAGGATTAATGAGGAGCTG 1560
Db 1501 TCTTACCAAAACCCAGGAGTACCCAGCTTATGAGAGATTTCTGAGGATTAATGAGGAGCTG 1560
Oy 1561 TGAAGTGAAGTG 1572
Db 1561 TGAAGTGAAGTG 1572
```

```
RESULT 2
US-10-342-887-775
Sequence 775, Application US/10342887
GENERAL INFORMATION:
APPLICANT: Dai, Hongyue
APPLICANT: He, Jidong
APPLICANT: Linsley, Peter S.
APPLICANT: Mao, Mao
APPLICANT: Roberts, Christopher J.
APPLICANT: Van 't Veer, Laura Johanna
APPLICANT: Van de Vijver, Marc J.
APPLICANT: Bernards, Rene
TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
FILE REFERENCE: 9301-188-999
CURRENT APPLICATION NUMBER: US/10/342,887
CURRENT FILING DATE: 2003-01-15
PRIOR APPLICATION NUMBER: 60/298,918
PRIOR FILING DATE: 2001-06-18
PRIOR FILING DATE: 2002-05-14
PRIOR FILING DATE: 2002-06-14
PRIOR FILING DATE: 2002-06-14
NUMBER OF SEQ ID NOS: 2699
SEQ ID NO 775
LENGTH: 1721
TYPE: DNA
ORGANISM: Homo sapiens
US-10-342-887-775
```

```
Query Match 94.2%; Score 1481.2; DB 9; Length 1721;
Best Local Similarity 95.5%; Pred. No. 31e-278;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;
```



```
Db 325 GAACAGCTTCAAGTTCAAGTCCACCTGAGGAGACAGATGTCACTTCGAGTCCAGTCAAC 384
QY 370 AGGCGAGCCCTTGGGCTTCCACACCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 429
Db 385 AGGCGAGCCCTTGGGCTTCCACACCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 444
QY 430 AAGCCAGCCCGGAGCTCCACCGCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 489
Db 445 AAGCCAGCCCGGAGCTCCACCGCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 474
QY 490 AAGCGAGCCCGGAGCTCCACCGCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 549
Db 475 -----GCCAAGGTGTCACTTCAGTCCCGGAGCCCAAG 504
QY 550 AAGCGAGCCCGGAGCTCCACCGCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 609
Db 505 AAGCGAGCCCGGAGCTCCACCGCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 564
QY 610 AAGCGAGCCCTTGGGCTTCCACCGCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 669
Db 565 AAGCGAGCCCTTGGGCTTCCACCGCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 624
QY 670 GCAATCAGGCTCAGCTTCTACTCTGTGTCAACCGGCACTTGTCCAGGCTTCCACAAC 729
Db 625 GCAATCAGGCTCAGCTTCTACTCTGTGTCAACCGGCACTTGTCCAGGCTTCCACAAC 684
QY 730 CCAGCCAGAGAGCACTCCACCCAGCATTTCCAGCCACCACTGTGATCTCTTCAAC 789
Db 685 CCAGCCAGAGAGCACTCCACCCAGCATTTCCAGCCACCACTGTGATCTCTTCAAC 744
QY 790 CTGGCAGGATAGCAACAGACTGATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 849
Db 745 CTGGCAGGATAGCAACAGACTGATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 804
QY 850 CTGACCTCTCTCAATCAGAGACTTCTCCGAGTGTCTACTGGGGTCTCTTCTTTTTC 909
Db 805 CTGACCTCTCTCAATCAGAGACTTCTCCGAGTGTCTACTGGGGTCTCTTCTTTTTC 864
QY 910 CTGCTTTTCAATTTCAACCTTCAGTTAATTTCTCTCTGGAAGATCCAGCAC 969
Db 865 CTGCTTTTCAATTTCAACCTTCAGTTAATTTCTCTCTGGAAGATCCAGCAC 924
QY 970 TACTACAGAGAGCTCAGAGAGCAATTTCTGAATGTTTTCAGATTTATTAACAAGG 1029
Db 925 TACTACAGAGAGCTCAGAGAGCAATTTCTGAATGTTTTCAGATTTATTAACAAGG 984
QY 1030 GGTTCCTGGGCTCTCAGATTAATTAAGTTCAGGCAAGATCTGTGGGTCAATTAAGT 1089
Db 985 GGTTCCTGGGCTCTCAGATTAATTAAGTTCAGGCAAGATCTGTGGGTCAATTAAGT 1044
QY 1090 CTGAGCTTCAGAGAGTACCATCAATGTCCAGAGCTGTGAGACAGATTCAATGAT 1149
Db 1045 CTGAGCTTCAGAGAGTACCATCAATGTCCAGAGCTGTGAGACAGATTCAATGAT 1104
QY 1150 AAAACGGAAGAGCTCTCAGATTAATCTGATGATCTCAGAGCTGAGGATGATG 1209
Db 1105 AAAACGGAAGAGCTCTCAGATTAATCTGATGATCTCAGAGCTGAGGATGATG 1164
QY 1210 CCAATTTCTTTTCTGTGCGCACTGTGGGGCTGTGGGGCTGTGGGGCTGTGGGGCTGTG 1269
Db 1165 CCAATTTCTTTTCTGTGCGCACTGTGGGGCTGTGGGGCTGTGGGGCTGTGGGGCTGTG 1224
QY 1270 GTGCTGTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1329
Db 1225 GTGCTGTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1284
QY 1330 CAGTGCAGGCGGAAAGAACTAGCGGAGAGTGAATCTTTCAGCGCGGAGTACCTACAT 1389
Db 1285 CAGTGCAGGCGGAAAGAACTAGCGGAGAGTGAATCTTTCAGCGCGGAGTACCTACAT 1344
QY 1390 CCAATGAGAGATACCTCACTTACCACTTACCACTTACCACTTACCACTTACCACTTAC 1449
Db 1345 CCAATGAGAGATACCTCACTTACCACTTACCACTTACCACTTACCACTTACCACTTAC 1404
```

```
QY 1450 GATGTAGCCCTTATGAGAGGTTTCTGAGGATATGTGTGACACACCTCTCTTACACA 1509
Db 1405 GATGTAGCCCTTATGAGAGGTTTCTGAGGATATGTGTGACACACCTCTCTTACACA 1464
QY 1510 AACCAGAGTGGAGAGCACTTCTGCAACTTGTAGAGGAGAGTGTGAGCTGTAGTGA 1569
Db 1465 AACCAGAGTGGAGAGCACTTCTGCAACTTGTAGAGGAGAGTGTGAGCTGTAGTGA 1524
QY 1570 TGG 1572
Db 1525 TGG 1527

RESULT 4
US-60-459-782-30
; Sequence 30, Application US/60459782
; GENERAL INFORMATION:
; APPLICANT: Wyeth Research
; APPLICANT: Twine, Natalie C.
; APPLICANT: Stover, Jennifer A.
; APPLICANT: Trepicchio, William L.
; APPLICANT: Dornier, Andrew J.
; APPLICANT: Stonim, Donna K.
; APPLICANT: Burczynski, Michael E.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND/OR SOLID TUMORS
; FILE REFERENCE: AM101055
; CURRENT APPLICATION NUMBER: US/60/459,782
; CURRENT FILING DATE: 2003-04-03
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 1804
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-459-782-30

Query Match 89.8%; Score 1412.2; DB 10; Length 1804;
Best Local Similarity 95.3%; Pred. No. 7.9e-265;
Matches 1490; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

QY 10 GAGCTGTTGATCTGTTCTGCGCCCTCCCAACCATTTTCAACACACCATGACACCGGAG 69
Db 25 GCGTGGTAATCTGTTCTGCGCCCTCCCAACCATTTTCAACACACCATGACACCGGAG 84
QY 70 ACCAGTCTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 129
Db 85 ACCAGTCTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 144
QY 130 GGTGATCAAGTCTTACCCAGAGTGAAGAAAGAGAGAGTGTGCTTACCCAGAGAGTTCA 189
Db 145 GGTGATCAAGTCTTACCCAGAGTGAAGAAAGAGAGAGTGTGCTTACCCAGAGAGTTCA 204
QY 190 GTGCGCAGCTTCTAGAGAGATGCTGATGATGACCAAGAGAGTACTCTCAGACAC 249
Db 205 GTGCGCAGCTTCTAGAGAGATGCTGATGATGACCAAGAGAGTACTCTCAGACAC 264
QY 250 AGCCCGGTTGAGGCTCTCCAGCACTCAGAGAGAGAGATGCTGCTGCGCCGAGAG 309
Db 265 AGCCCGGTTGAGGCTCTCCAGCACTCAGAGAGAGAGATGCTGCTGCGCCGAGAG 324
QY 310 GAACAGCTTCAAGTTCAAGTGTGCACTGTGGAGACAGATGTCACTTGTGTCCAGTACC 369
Db 325 GAACAGCTTCAAGTTCAAGTGTGCACTGTGGAGACAGATGTCACTTGTGTCCAGTACC 384
QY 370 AAGCCAGGCTTGGGCTTCCACACCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 429
Db 385 AAGCCAGGCTTGGGCTTCCACACCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 444
QY 430 AAGCCAGGCTTGGGCTTCCACACCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 489
Db 445 AAGCCAGGCTTGGGCTTCCACACCCCGGAGCCCAAGATGTCACTTCAGTCCCGGACAAC 474
```



```

670 GCATCAGGCTCAGGCTTCTACTCTGTGTCACAAAGGCACTTGGCAGGGCTTACCAAGC 729
625 GATTCAGGCTCAGGCTTCTACTCTGTGTCACAAAGGCACTTGGCAGGGCTTACCAAGC 684
730 CCAGCCAGAGAGAGACTTCCACCCAGCTTCCAGCCACCACTTGATCTCTTACCAACC 789
685 CCAGCCAGAGAGAGACTTCCACCCAGCTTCCAGCCACCACTTGATCTCTTACCAACC 744
790 CTTCAGCAGCAGTACCAACCAAGCTGATGCACTGACCACTTACCAAGCAGTACCTCT 849
745 CTTCAGCAGCAGTACCAACCAAGCTGATGCACTGACCACTTACCAAGCAGTACCTCT 804
850 CTTCAGCAGCAGTACCAACCAAGCTGATGCACTGACCACTTACCAAGCAGTACCTCT 909
805 CTTCAGCAGCAGTACCAACCAAGCTGATGCACTGACCACTTACCAAGCAGTACCTCT 864
910 CTTCAGCAGCAGTACCAACCAAGCTGATGCACTGACCACTTACCAAGCAGTACCTCT 969
865 CTTCAGCAGCAGTACCAACCAAGCTGATGCACTGACCACTTACCAAGCAGTACCTCT 924
970 TACTACCAAGAGCTGCAAGAGCACTTCTGAATGTTTTCGACATTTTAAACAAGG 1029
925 TACTACCAAGAGCTGCAAGAGCACTTCTGAATGTTTTCGACATTTTAAACAAGG 984
1030 GGTTCCTGAGGCTCTCTCAATATTAATTAAGTTCAGGCTGAGATCTGTGTGTGTA 1089
985 GGTTCCTGAGGCTCTCTCAATATTAATTAAGTTCAGGCTGAGATCTGTGTGTGTA 1044
1090 CTGGCTCTCCAGAGAGTACCATCATGTCACGACGTGAGACAGAGTACCATCATGAT 1149
1045 CTGGCTCTCCAGAGAGTACCATCATGTCACGACGTGAGACAGAGTACCATCATGAT 1104
1150 AAAACGAGAGAGCTCTGATATTAACCTGACGATCTCAGACGTGACGATGATGATG 1209
1105 AAAACGAGAGAGCTCTGATATTAACCTGACGATCTCAGACGTGACGATGATGATG 1164
1210 CCAATTCCTCTCTCTGACGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1269
1165 CCAATTCCTCTCTCTGACGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1224
1270 GTGCTGATCTGTGTTGTGTTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1329
1225 GTGCTGATCTGTGTTGTGTTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1284
1330 CAGTGCCTCCGAGAGAGTACGAGGAGCTGAGCATCTTTCAGCCCGGAGATACCTACAT 1389
1285 CAGTGCCTCCGAGAGAGTACGAGGAGCTGAGCATCTTTCAGCCCGGAGATACCTACAT 1344
1390 CCTATGAGGAGTACCCCACTTACCAACCCATGAGGCTGAGGCTGAGGCTGAGGCTGAG 1449
1345 CCTATGAGGAGTACCCCACTTACCAACCCATGAGGCTGAGGCTGAGGCTGAGGCTGAG 1404
1450 GATCGTACGCTTATGAGAGGTTCTGAGGTAATGTTGAGGAGGCTGAGGCTGAGGCTGAG 1509
1405 GATCGTACGCTTATGAGAGGTTCTGAGGTAATGTTGAGGAGGCTGAGGCTGAGGCTGAG 1464
1510 AACCCAGAGAGGAGGCTTCTGCACTTGTAGGAGGAGCTGAGGCTTGTAGGCTGAG 1569
1465 AACCCAGAGAGGAGGCTTCTGCACTTGTAGGAGGAGCTGAGGCTTGTAGGCTGAG 1524
1570 TGG 1572
1525 TGG 1527
```

```

CURRENT APPLICATION NUMBER: US/10/101,510
CURRENT FILING DATE: 2002-03-20
PRIOR APPLICATION NUMBER: 60/276,947
FRIOR FILING DATE: 2001-03-20
NUMBER OF SEQ ID NOS: 805
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 339
LENGTH: 1823
TYPE: DNA
ORGANISM: Homo sapiens
US-10-101-510-339

Query Match      89.8%; Score 1412.2; DB 9; Length 1823;
Best Local Similarity 95.3%; Pred. No. 7.9e-265;
Matches 1490; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

10 GGTGCTGTAATCTGTTCTGTCGCTCCCTCCCATTTGACCAACCAAGAGAGAGGAG 69
25 GCTGCTGTAATCTGTTCTGTCGCTCCCTCCCATTTGACCAACCAAGAGAGAGGAG 84
70 ACCGATCTCTCTTCTTCTGCTGCTGCTCTCAAGTCTTACATGTTTGAAGTTCT 129
85 ACCGATCTCTCTTCTTCTGCTGCTGCTCTCAAGTCTTACATGTTTGAAGTTCT 144
130 GGTCAATGCAAGCTTACCCAGAGTGAAGAAAGAGACTTCGGCTTACCAAGAGTTCA 189
145 GGTCAATGCAAGCTTACCCAGAGTGAAGAAAGAGACTTCGGCTTACCAAGAGTTCA 204
190 GTGGCCAGCTTACTGAGAAAGTCTGTGATGATGACAGAGAGGATCTCTCCAGCCAC 249
205 GTGGCCAGCTTACTGAGAAAGTCTGTGATGATGACAGAGAGGATCTCTCCAGCCAC 264
250 AGCCCGGATGAGGCTCTCTCCACCACTGAGGAGACAGATGATGATCTGAGGAGGAG 309
265 AGCCCGGATGAGGCTCTCTCCACCACTGAGGAGACAGATGATGATCTGAGGAGGAG 324
310 GAAACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 369
325 GAAACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 384
370 AAGCCAGGCTCGGAGCTTCCACCAAGCCCGAGCCAGAGTATGATGATGATGATGATG 429
385 AAGCCAGGCTCGGAGCTTCCACCAAGCCCGAGCCAGAGTATGATGATGATGATGATG 444
430 AAGCCAGGCTCGGAGCTTCCACCAAGCCCGAGCCAGAGTATGATGATGATGATGATG 489
445 AAGCCAGGCTCGGAGCTTCCACCAAGCCCGAGCCAGAGTATGATGATGATGATGATG 474
490 AAGCCAGGCTCGGAGCTTCCACCAAGCCCGAGCCAGAGTATGATGATGATGATGATG 549
475 -----GCCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 504
550 AAGCCAGGCTCGGAGCTTCCACCAAGCCCGAGCCAGAGTATGATGATGATGATGATG 609
505 AAGCCAGGCTCGGAGCTTCCACCAAGCCCGAGCCAGAGTATGATGATGATGATGATG 564
610 AAGCCAGGCTCGGAGCTTCCACCAAGCCCGAGCCAGAGTATGATGATGATGATGATG 669
565 AAGCCAGGCTCGGAGCTTCCACCAAGCCCGAGCCAGAGTATGATGATGATGATGATG 624
670 GATCAGGCTCAGCTTCTACTCTGTGTCACAAAGGAGCTTGTGAGGAGGCTTACCAAGC 729
625 GATCAGGCTCAGCTTCTACTCTGTGTCACAAAGGAGCTTGTGAGGAGGCTTACCAAGC 684
730 CCAGCCAGAGAGAGCTTCCACCCAGCTTCCAGCCACCACTTGATCTCTTACCAACC 789
685 CCAGCCAGAGAGAGCTTCCACCCAGCTTCCAGCCACCACTTGATCTCTTACCAACC 744
790 CTTCAGCAGCAGTACCAACCAAGCTGATGCACTGACCACTTACCAAGCAGTACCTCT 849
745 CTTCAGCAGCAGTACCAACCAAGCTGATGCACTGACCACTTACCAAGCAGTACCTCT 804
850 CTTCAGCAGCAGTACCAACCAAGCTGATGCACTGACCACTTACCAAGCAGTACCTCT 909
```


Db	1021	GGCTTTCCGAGAAAGGTACATCAATCATTCACACGAGCTGGAGACACAGTTCAATCAAGTATTA	1080
Db	1152	AAGGAAAGAGGCTCTGGATTTAACTTGAACATTCAGACGTCAAGCTGAAGTCATGTGCC	1211
Db	1081	AAGGAAAGAGGCTCTGCAATTAATTAACCTGACAGATTCAGACGTCAAGCTGAAGTATGTGCC	1140
Qy	1212	ATTTCTTTCTCTGCCACAGTCTGGGGCTGGGGTGGCCAGGCTGGGGAGCATGCGCTGTGGT	1271
Db	1141	ATTTCTTTCTCTGCCAGTCTGGGGCTGGGGTGGCCAGGCTGGGGAGCATGCGCTGTGGT	1200
Qy	1272	GCTGTCTGTGTTCTGGTTGGCGTGGCCATTTGTATCTCATTTGGCTTGGCGTGTGTGCA	1331
Db	1201	GCTGTCTGTGTTCTGGTTGGCTGGCCATTTGTATCTCATTTGGCTTGGCTGTGTGTGCA	1260
Qy	1332	GTTCCCGCGGAAAGAACTACGGGAGCTGTGACATCTTTTCAGCCCGGGAATACCTTACATCC	1391
Db	1261	GTTCCCGCGGAAAGAACTACGGGAGCTGTGACATCTTTTCAGCCCGGGAATACCTTACATCC	1320
Qy	1392	TATGACCGAGTACCCACCTTACACACCCCATGGGCGCTATGTGCCCCCTTAGCAGTACCGA	1451
Db	1321	TATGACCGAGTACCCACCTTACACACCCCATGGGCGCTATGTGCCCCCTTAGCAGTACCGA	1380
Qy	1452	TGCTTACCCCTTATGAGAAAGTTTCTGACAGTAAATGTGGGACAGAGCTCTCTTTCACAAA	1511
Db	1381	TGCTTACCCCTTATGAGAAAGTTTCTGACAGTAAATGTGGGACAGAGCTCTCTTTCACAAA	1440
Qy	1512	CCGACAGATGGGCACACTTCTGTGCACATTGTAGGGGACAGTGGCCCTTGTAGCTGAGTG	1571
Db	1441	CCGACAGATGGGCACACTTCTGTGCACATTGTAGGGGACAGTGGCCCTTGTAGCTGAGTG	1500
Qy	1572		
Db	1501		

RESULT 8
US-60-452-680-471
Sequence 471, Application US/60452680
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: C1001450
CURRENT APPLICATION NUMBER: US/60/452,680
NUMBER OF SEQ ID NOS: 116213
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 471
LENGTH: 1549
TYPE: DNA
ORGANISM: Homo sapiens
US-60-452-680-471

Query Match 89.7%; Score 1409.8; DB 11; Length 1549;
Best Local Similarity 95.1%; Pred. No. 2.3e-264;
Matches 1484; Conservative 5; Mismatches 12; Indels 60; Gaps 1;

Qy	12	CTGCTTGAAATCTGTTTGGCCCCCTCCCCACCATTTTCAACCAACATATGACACCGGGCAC	71
Db	1	CTGCTGGAATCTGTTTGGCCCCCTCCCCACCATTTTCAACCAACATATGACACCGGGCAC	60
Qy	72	CCAGTCTCCTTCTTCTCGTGCCTGCTCTCTCAAGTGCCTTACAGTTTACAGGTTCTGG	131
Db	61	CCAGTCTCCTTCTTCTCGTGCCTGCTCTCTCAAGTGCCTTACAGTTTACAGGTTCTGG	120
Qy	132	TATGCAAGCTCTTACCCAGGTGGAATAAAGGACACTTGGGCTACCCAGAGAAATTCAGT	191
Db	121	TATGCAAGCTCTTACCCAGGTGGAATAAAGGACACTTGGGCTACCCAGAGAAATTCAGT	180
Qy	192	GCCCAAGCTCTTACGAAAGATGCTGTGATATGACAGACGTATCTTCCAGCCACAG	251

D	181	GCACGCTCACTGAGAAAGATGCTGTGAGTATGACACGAGCGTATCTTCACGACACAG	240
O	252	CCCCCGTTGAGGCTCTCCACCACTCAGGGGACAGGATGTCACTTGGCCCCGACACGGA	311
D	241	CCCCGGTTCAGGCTCTCCACCACTCAGGGGACAGGATGTCACTTGGCCCCGACACGGA	300
O	312	ACCAGCTTCAGTTCAGCTGACCTTGGGGAACGAGATGTCACTCGGATCCCACTACAG	371
D	301	ACCAGCTTCAGTTCAGCTGACCTTGGGGAACGAGATGTCACTCGGATCCCACTACAG	360
O	372	GCACGCTCTGGGCTCCACACCCCGGACGCGGACGATGTACTGACCTGACGCCCGGACAA	431
D	361	GCACGCTCTGGGCTCCACACCCCGGACGCGGACGATGTACTGACCTGACGCCCGGACAA	420
O	432	GCACGCTCTGGGCTCCACCGCCCTCCCGGACGATGTCACTTGGCCCCGACACAG	491
D	421	GCACGCTCTGGGCTCCACCGCCCTCCCGGACGATGTCACTTGGCCCCGACACAG	448
O	492	GCACGCTCTGGGCTCCACCGCCCGGACGATGTCACTTGGCCCCGACACAG	551
D	449	-----GCCACGATGTCACTTGGCCCCGACACAG	480
O	552	GCACGCTCTGGGCTCCACCGCCCGGACGATGTCACTTGGCCCCGACACAG	611
D	481	GCACGCTCTGGGCTCCACCGCCCGGACGATGTCACTTGGCCCCGACACAG	540
O	612	GCACGCTCTGGGCTCCACCGCCCTCCCGGACGATGTCACTTGGCCCCGACACAG	671
D	541	GCACGCTCTGGGCTCCACCGCCCTCCCGGACGATGTCACTTGGCCCCGACACAG	600
O	672	ATCAGGCTCAGCTTCACTGTGTGACCAACGAGCCTGTGACGAGGCTTACACACCCG	731
D	601	ATCAGGCTCAGCTTCACTGTGTGACCAACGAGCCTGTGACGAGGCTTACACACCCG	660
O	732	AGCAGCAAGAGACCTCAACCCGACATTCGACGACCACTCGATGATCTCGTACACCT	791
D	661	AGCAGCAAGAGACCTCAATTCGATTCACATTCGACGACCACTCGATGATCTCGTACACCT	720
O	792	TGCGACCATAGACCAAGACATGTGCGAGTACCTACCATGACACGATCTCTCTCT	851
D	721	TGCGACCATAGACCAAGACATGTGCGAGTACCTACCATGACACGATCTCTCTCT	780
O	852	CACGCTCTCAATCACACGACTTCTCCGAGTGTCACTGGGGGCTCTTTCTTTTCT	911
D	781	CACGCTCTCAATCACAGGACTTCTCCGAGTGTCACTGGGGGCTCTTTCTTTTCT	840
O	912	GCTCTTTCACATTTAAACCTCAGATTAAATTCCTCTGTGAAAGATCCGACACGACAT	971
D	841	GCTCTTTCACATTTAAACCTCAGATTAAATTCCTCTGTGAAAGATCCGACACGACAT	900
O	972	CTACCAAGACTCTGACAGACGATTTCTGAAATGTTTTGACGATTTTAAACAAGGGG	1031
D	901	CTACCAAGACTCTGACAGACGATTTCTGAAATGTTTTGACGATTTTAAACAAGGGG	960
O	1032	TTTTCTGGGCTCTCCAAATTTAGTTGAGGCGAGATCTGTGGTGTGATTAAGACTCT	1091
D	961	TTTTCTGGGCTCTCCAAATTTAGTTGAGGCGAGATCTGTGGTGTGATTAAGACTCT	1020
O	1092	GGCTCTTCGAGAGGTAACCATCAATGTCCAGAGGTGTGAGACACGTTCAATCAGAT	1151
D	1021	GGCTCTTCGAGAGGTAACCATCAATGTCCAGAGGTGTGAGACACGTTCAATCAGAT	1080
O	1152	AACGAGACGACTCTGATATACTGACGATCTCAGAGCTCAGACGTCGATCATGTGCT	1211
D	1081	AACGAGACGACTCTGATATACTGACGATCTCAGAGCTCAGAGCTCAGATCATGTGCT	1140
O	1212	ATTTCCTTTCTCGCCGACGTCTGGGGCTGGGGGTGCGACGCTGGGGACACCGCGCTGCT	1271
D	1141	ATTTCCTTTCTCGCCGACGTCTGGGGCTGGGGGTGCGACGCTGGGGACACCGCGCTGCT	1200
O	1272	GCTGGTCTGTGTTCTGTTTGCGTGTGCATTTGTCTATCTCATTTGCTTGGCTGTCTCA	1331
D	1201	GCTGGTCTGTGTTCTGTTTGCGTGTGCATTTGTCTATCTCATTTGCTTGGCTGTCTCA	1260

Oy 1332 GTGCGCGCGAAGAACTACGCGGAGCTGACATCTTTCCAGCCCGGGATACACCATCC 1391
 Db 1361 GTGCGCGCGAAGAACTACGCGGAGCTGACATCTTTCCAGCCCGGGATACACCATCC 1320
 Oy 1392 TATGAGCGAGTACCCCACTTACACACACACACACACACACACACACACACACAC 1451
 Db 1321 TATGAGCGAGTACCCCACTTACACACACACACACACACACACACACACACACAC 1380
 Oy 1452 TCGTAGACCCCTATGAGAGTCTTCTGAGGATATGAGGAGAGAGAGAGAGAGAGAG 1511
 Db 1381 TCGTAGACCCCTATGAGAGTCTTCTGAGGATATGAGGAGAGAGAGAGAGAGAGAG 1440
 Oy 1512 CCGAG 1571
 Db 1441 CCGAG 1500
 Oy 1572 G 1572
 Db 1501 G 1501

RESULT 9
 US-60-453-135-343
 ; Sequence 343, Application US/60453135
 ; GENERAL INFORMATION:
 ; APPLICANT: CARILL, Michele
 ; APPLICANT: IAKUBOVA, Olga
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; FILE REFERENCE: CL001456
 ; CURRENT APPLICATION NUMBER: US/60/453,135
 ; CURRENT FILING DATE: 2003-03-10
 ; NUMBER OF SEQ ID NOS: 82762
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 343
 ; LENGTH: 1549
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-60-453-135-343

Query Match 89.7%; Score 1409.8; DB 11; Length 1549;
 Best Local Similarity 95.1%; Pred. No. 2.3e-264;
 Matches 148; Conservative 5; Mismatches 12; Indels 60; Gaps 1;

Oy 12 CTGCTGAACTCTCTGCGCCCTCCCACTTTCACACACACACACACACACACACAC 71
 Db 1 CTGCTGAACTCTCTGCGCCCTCCCACTTTCACACACACACACACACACACACAC 60
 Oy 72 CCACTCTCTTCTCTGCTGCTGCTCTCTACAGTCTTACAGTTCTTACAGTTCTG 131
 Db 61 CCACTCTCTTCTCTGCTGCTGCTCTCTACAGTCTTACAGTTCTTACAGTTCTG 120
 Oy 132 TCATGCAAGCTTACCCAGAGTGAAGAAAGAGACTTCGGCTACCCAGAGAGTTCA 191
 Db 121 TCATGCAAGCTTACCCAGAGTGAAGAAAGAGACTTCGGCTACCCAGAGAGTTCA 180
 Oy 192 GCCAGCTTACCTGAGAGAAATGCTGATGATGACACACACAGTCTCTCAGCCACAG 251
 Db 181 GCCAGCTTACCTGAGAGAAATGCTGATGATGACACACACAGTCTCTCAGCCACAG 240
 Oy 252 CCCCAGTTTCAAGGCTCTCTCACTCACTCAAGGACAGAGATGCTACTGAGCCCGGCA 311
 Db 241 CCCCAGTTTCAAGGCTCTCTCACTCACTCAAGGACAGAGATGCTACTGAGCCCGGCA 300
 Oy 312 ACCAGCTTCAAGTTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 371
 Db 301 ACCAGCTTCAAGTTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 360
 Oy 372 GCCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 431
 Db 361 GCCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 420

Oy 432 GCCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 491
 Db 421 GCCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 448
 Oy 492 GCCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 551
 Db 449 -----GCCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 480
 Oy 552 GCCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 611
 Db 481 GCCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 540
 Oy 612 GCCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 671
 Db 541 GCCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 600
 Oy 672 ATCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 731
 Db 601 ATCAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 660
 Oy 732 AGCAGCAAGAGCACTTCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 791
 Db 661 AGCAGCAAGAGCACTTCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 720
 Oy 792 TGCAGAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 851
 Db 721 TGCAGAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 780
 Oy 852 CACCTCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 911
 Db 781 CACCTCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 840
 Oy 912 GTCTTTCATTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 971
 Db 841 GTCTTTCATTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 900
 Oy 972 CTACAGAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 1031
 Db 901 CTACAGAGCTTCAAGTCCCACTGAGGACAGAGATGCTACTGAGTTCCAGTCA 960
 Oy 1032 TTTTCTGAGGCTTCTCAATTTAAGTTCAAGGACAGAGATGCTACTGAGTTCCAGTCA 1091
 Db 961 TTTTCTGAGGCTTCTCAATTTAAGTTCAAGGACAGAGATGCTACTGAGTTCCAGTCA 1020
 Oy 1092 GGCCTTCCAGAGAGTCAATTCATATGTCACAGAGTGAAGACAGAGTCAATGATTA 1151
 Db 1021 GGCCTTCCAGAGAGTCAATTCATATGTCACAGAGTGAAGACAGAGTCAATGATTA 1080
 Oy 1152 AACGAGAGAGCTTCTGATTTAACTGACATCTTCAAGTCAAGTCAAGTCAAGTCA 1211
 Db 1081 AACGAGAGAGCTTCTGATTTAACTGACATCTTCAAGTCAAGTCAAGTCAAGTCA 1140
 Oy 1212 ATTTCTTCTCTGCGCCAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1271
 Db 1141 ATTTCTTCTCTGCGCCAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1200
 Oy 1272 GCTGATCTGATTTCTGATTTGAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1331
 Db 1201 GCTGATCTGATTTCTGATTTGAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1260
 Oy 1332 GTGCGCGAAGAACTACGCGGAGTGAAGTCTTTCCAGCCCGGGATACACATCC 1391
 Db 1261 GTGCGCGAAGAACTACGCGGAGTGAAGTCTTTCCAGCCCGGGATACACATCC 1320
 Oy 1392 TATGAGCGAGTACCCCACTTACACACACACACACACACACACACACACACACAC 1451
 Db 1321 TATGAGCGAGTACCCCACTTACACACACACACACACACACACACACACACACAC 1380
 Oy 1452 TCGTAGACCCCTATGAGAGTCTTCTGAGGATATGAGGAGAGAGAGAGAGAGAGAG 1511
 Db 1381 TCGTAGACCCCTATGAGAGTCTTCTGAGGATATGAGGAGAGAGAGAGAGAGAGAG 1440
 Oy 1512 CCGAG 1571


```

; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038PC
; CURRENT APPLICATION NUMBER: PCT/US02/19669A
; CURRENT FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 310
; LENGTH: 4139
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-19669A-310

```

```

Query Match      75.7%; Score 1190.2; DB 1; Length 4139;
Best Local Similarity 94.8%; Pred. No. 1.2e-221;
Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

```

Oy 249 CAGCCCGGTTAGGCTCTCCACACTCAGGAGACAGATGTCATCTGGCCCCGAC 308
Db 2545 CAGCCCGGTTAGGCTCTCCACACTCAGGAGACAGATGTCATCTGGCCCCGAC 2604
Oy 309 GGAACAGAGTTAGGTTGAGTGCACCTGAGGAGAGATGTCATCTGGTTCCAGTTC 368
Db 2605 CAGGCGGCGCGCGGCTCCACCGCGCCCGACGCGTGTACCTCGGCCGAGACAC 2664
Oy 369 CAGGCGGCGCGGCTCCACCGCGCCCGACGCGTGTACCTCGGCCGAGACAC 428
Db 2665 CAGGCGGCGCGGCTCCACCGCGCCCGACGCGTGTACCTCGGCCGAGACAC 2724
Oy 429 CAGGCGGCGCGGCTCCACCGCGCCCGACGCGTGTACCTCGGCCGAGACAC 488
Db 2725 CAGGCGGCGCGGCTCCACCGCGCCCGACGCGTGTACCTCGGCCGAGACAC 2784
Oy 489 CAGGCGGCGCGGCTCCACCGCGCCCGACGCGTGTACCTCGGCCGAGACAC 548
Db 2785 CAGGCGGCGCGGCTCCACCGCGCCCGACGCGTGTACCTCGGCCGAGACAC 2844
Oy 549 CAGGCGGCGCGGCTCCACCGCGCCCGACGCGTGTACCTCGGCCGAGACAC 608
Db 2845 CAGGCGGCGCGGCTCCACCGCGCCCGACGCGTGTACCTCGGCCGAGACAC 2904
Oy 609 CAGGCGGCGCGGCTCCACCGCGCCCGACGCGTGTACCTCGGCCGAGACAC 668
Db 2905 CAGGCGGCGCGGCTCCACCGCGCCCGACGCGTGTACCTCGGCCGAGACAC 2964
Oy 669 TGCAATAGAGCTGAGTTGATCTGAGTGCACAGGCACTCTGCGAGGCTACACAC 728
Db 2965 TGCAATAGAGCTGAGTTGATCTGAGTGCACAGGCACTCTGCGAGGCTACACAC 3024
Oy 729 CCAAGCCAGAGAGACTCACAACCAAGATTCCAGCCACCACTCTGATACCTCAAC 788
Db 3025 CCAAGCCAGAGAGACTCACAACCAAGATTCCAGCCACCACTCTGATACCTCAAC 3084
Oy 789 CATTGCAAGCAATAGCAACAGATGTCAGTACCTCAACCTGACAGGATCACTCC 848
Db 3085 CATTGCAAGCAATAGCAACAGATGTCAGTACCTCAACCTGACAGGATCACTCC 3144
Oy 849 TCTCACTCTCTCAATCAGACACTCTCTCCCGAGTTGTCTACTGAGGCTCTTTTCTTTT 908
Db 3145 TCTCACTCTCTCAATCAGACACTCTCTCCCGAGTTGTCTACTGAGGCTCTTTTCTTTT 3204
Oy 909 CTTGTCTTTTCAATTGAAACCTCCAGTTTAATCTCTCTGAGAGATCCAGACCGA 968

```

```

Db 3205 CTTGTCTTTTCAATTGAAACCTCCAGTTTAATCTCTCTGAGAGATCCAGACCGA 3264
Oy 969 CTACTTCCAGAGCTGCAGAGAGACATTTCTGAATGTTTTGAGATTAATAACAAG 1028
Db 3265 CTACTTCCAGAGCTGCAGAGAGACATTTCTGAATGTTTTGAGATTAATAACAAG 3324
Oy 1029 GGGTTTCTGGGCTCTCCAAATTAAGTTCAAGGCAAGATCTGTGTGATCAATTGAC 1088
Db 3325 GGGTTTCTGGGCTCTCCAAATTAAGTTCAAGGCAAGATCTGTGTGATCAATTGAC 3384
Oy 1089 TCTGGCTTCCAGAGAGTACCATCATATGTCACAGACGTGAGAGACAGTTCAATCAGTA 1148
Db 3385 TCTGGCTTCCAGAGAGTACCATCATATGTCACAGACGTGAGAGACAGTTCAATCAGTA 3444
Oy 1149 TAAACGAGAGAGCTCTGATTAATTAAGTTCAAGTCAAGTCAAGTCAAGTCAAGT 1208
Db 3445 TAAACGAGAGAGCTCTGATTAATTAAGTTCAAGTCAAGTCAAGTCAAGTCAAGT 3504
Oy 1209 GCCATTCTCTTCTCTGCGCAATCTGGGCTGGGCTGGGCTGGGCTGGGCTGGT 1268
Db 3505 GCCATTCTCTTCTCTGCGCAATCTGGGCTGGGCTGGGCTGGGCTGGGCTGGT 3564
Oy 1269 GGTGCTGATCTGTGTTGAGTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGT 1328
Db 3565 GGTGCTGATCTGTGTTGAGTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGT 3624
Oy 1329 TCAATGCGCGCGAAGAACTACAGGAGAGTGTCTTTCAGCCCGGATACCTACCA 1388
Db 3625 TCAATGCGCGCGAAGAACTACAGGAGAGTGTCTTTCAGCCCGGATACCTACCA 3684
Oy 1389 TCCATGAGCAGATACCCCACTACACACACCCATGAGGCGCTATGCCCCCTAGAGTAC 1448
Db 3685 TCCATGAGCAGATACCCCACTACACACACCCATGAGGCGCTATGCCCCCTAGAGTAC 3744
Oy 1449 CGATGTCAGCCCTATGAGAAAGTTCTGCAAGTAAATGTTGAGCAGAGCTCTTACAC 1508
Db 3745 CGATGTCAGCCCTATGAGAAAGTTCTGCAAGTAAATGTTGAGCAGAGCTCTTACAC 3804
Oy 1509 AAACCCAGAGTGGCAGACACTTCTGCAACTTTGTAGGG 1547
Db 3805 AAACCCAGAGTGGCAGACCGCTTCTGCAACTTTGTAGGG 3843

```

```

RESULT 12
US-10-170-235-24364
Sequence 24364, Application US/10170235
GENERAL INFORMATION:

```

```

; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF HUMAN
; FILE REFERENCE: C1001380
; CURRENT APPLICATION NUMBER: US/10/170,235
; CURRENT FILING DATE: 2003-03-17
; NUMBER OF SEQ ID NOS: 42514
; SEQ ID NO 24364
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: HUMAN
US-10-170-235-24364

```

```

Query Match      60.7%; Score 953.8; DB 8; Length 1353;
Best Local Similarity 81.0%; Pred. No. 8.1e-176;
Matches 1266; Conservative 0; Mismatches 12; Indels 285; Gaps 2;

```

```

Oy 10 GGGTGTGTAATCTGTTCTGCCCCCTCCACCACTTTTCAACACACACAGACCGGGC 69
Db 19 GCTGCTGAATCTGTTCTGCCCCCTCCACCACTTTTCAACACACACAGACCGGGC 78
Oy 70 ACCCAATCTCTTCTCTGCTGCTGCTCTCAAGTGTACAGTTGTTACAGTTCT 129
Db 79 ACCCAATCTCTTCTCTGCTGCTGCTCTCAAGTGTACAGTTGTTACAGTTCT 138

```



```

Db 319 GAACCAAGCTTCAAGTTCACTGCTGCACTGGGGACAAGATTGTCACCTGGTCCAGTACACC 378
Qy 370 AGGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 429
Db 379 AGGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 438
Qy 430 AAGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 489
Db 439 AAGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 466
Qy 490 AGGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 549
Db 467 -----CAGCCCAAGGTTGTCACCTGAGCCCGGACAAC 498
Qy 550 AGGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 609
Db 499 AGGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTGAGCCCGGACAAC 558
Qy 610 AAGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 669
Db 559 AAGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 618
Qy 670 GCATCAGGCTCAGCTTCTACTCTGGTGCACAACGAGACCTGCGCAGGCGTACCAACACC 729
Db 619 GCATCAGGCTCAGCTTCTACTCTGGTGCACAACGAGACCTGCGCAGGCGTACCAACACC 678
Qy 730 CGAGCGAGGAGAGAGATCCACCGAGATTCGCGAGCACTGATGATCTCTTACCAACC 789
Db 679 CGAGCGAGGAGAGATCCACCGAGATTCGCGAGCACTGATGATCTCTTACCAACC 738
Qy 790 CTTGCGAGCCATAGACCAAGAGCTGATGCAAGTACCACTCAACATAGACGAGTACCTCT 849
Db 739 CTTGCGAGCCATAGACCAAGAGCTGATGCAAGTACCACTCAACATAGACGAGTACCTCT 798
Qy 850 CTGACCTCTCAATCAGACGACTTCTCCAGATGTCATCACTGAGGAGTCTCTTCTTTTC 909
Db 799 CTGACCTCTCAATCAGACGACTTCTCCAGATGTCATCACTGAGGAGTCTCTTCTTTTC 858
Qy 910 CTGCTTTTCAATTTCAACCTCAAGTAACTCTCTCTGGAAGATCCGACACCGAC 969
Db 859 CTGCTTTTCAATTTCAACCTCAAGTAACTCTCTCTGGAAGATCCGACACCGAC 918
Qy 970 TACTACCAAGAGCTGAGAGACATTTCTGAATGTTTGGCAGATTTTAAACAAGAG 1029
Db 919 TACTACCAAGAGCTGAGAGACATTTCTGAAT----- 953
Qy 1030 GGTTCCTGGGCTCTCAATATTAAATTGAGGCCAGATCTGTGTGTACATTTGACT 1089
Db 954 ----- 953
Qy 1090 CTGGCTTCGAGAGAGTACCATATGTCACGAGCGTGGAGACAAGTTCAATCAGTAT 1149
Db 954 ----- 953
Qy 1150 AAAAGGAGAGAGCTCTGATATTAATTGAGATCTCAAGCTGAGGAGTATGAG 1209
Db 954 ----- 953
Qy 1210 CCAATTTCTTTTCTGCCCAGCTGAGGAGCTGGGAGTGGAGCTGAGGAGCTG 1269
Db 954 -----GTCTGGGAGCTGGGAGTGGAGCTGAGGAGTGGAGCTGAGGAGCTG 993
Qy 1270 GTGCTGAGTCTGTTCTGAGTTGGCTGGAGTATGTCATCTCATTTGAGCTGTGT 1329
Db 994 GTGCTGAGTCTGTTCTGAGTTGGCTGGAGTATGTCATCTCATTTGAGCTGTGT 1053
Qy 1330 CAGTGGCGCGGAGGAGGAGTATGAGGAGTATGTCATCTTCCAGCCCGGAGTATCACT 1389
Db 1054 CAGTGGCGCGGAGGAGGAGTATGAGGAGTATGTCATCTTCCAGCCCGGAGTATCACT 1113
Qy 1390 CTTATAGGAGGAGTATCCCACTTACCAAGCCATGAGGAGTATGTCATCTTCCAGTAC 1449
Db 1114 CTTATAGGAGGAGTATCCCACTTACCAAGCCATGAGGAGTATGTCATCTTCCAGTAC 1173

```

```

Qy 1450 GATGTAGCCCTTATGAGAGTTTCTGCAAGTATGTTGGAGACAGCTCTCTTACACA 1509
Db 1174 GATGTAGCCCTTATGAGAGTTTCTGCAAGTATGTTGGAGACAGCTCTCTTACACA 1233
Qy 1510 AACCCAGAGTGGAGAGCACTTCTGCACTTTGAGGGGACAGTGGCTTGGAGTGG 1569
Db 1234 AACCCAGAGTGGAGAGCACTTCTGCACTTTGAGGGGACAGTGGCTTGGAGTGG 1293
Qy 1570 TGG 1572
Db 1294 TGG 1296

RESULT 14
US-60-453-135-346
; Sequence 346, Application us/60453135
; GENERAL INFORMATION:
; APPLICANT: CARBELL, Michele
; APPLICANT: IAKOUBOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001456
; CURRENT APPLICATION NUMBER: US/60/453,135
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 346
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-453-135-346

Query Match 60.6%; Score 952.2; DB 11; Length 1353;
Best Local Similarity 80.7%; Pred. No. 1,6e-175;
Matches 1262; Conservative 4; Mismatches 12; Indels 285; Gaps 2;

Qy 10 GGTGCTTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 69
Db 19 GCTGCTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 78
Qy 70 ACCAGTCTCTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 129
Db 79 ACCAGTCTCTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 138
Qy 130 GGTGATGAGCTTACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 189
Db 139 GGTGATGAGCTTACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 198
Qy 190 GTGCCAGCTTCTAGAGAGATGCTGTGATGATGACCCAGAGCTTCTTCAAGCTAC 249
Db 199 GTGCCAGCTTCTAGAGAGATGCTGTGATGATGACCCAGAGCTTCTTCAAGCTAC 258
Qy 250 AGCCCGGTTGAGCTCTCTCACTCAAGGAGAGAGATGCTCTGAGCCCGGAGCAG 309
Db 259 AGCCCGGTTGAGCTCTCTCACTCAAGGAGAGAGATGCTCTGAGCCCGGAGCAG 318
Qy 310 GAACCAAGCTTCAAGTTCAAGTTCACCTGAGGAGAGAGATGTCACCTTGGTCCAGTAC 369
Db 319 GAACCAAGCTTCAAGTTCAAGTTCACCTGAGGAGAGAGATGTCACCTTGGTCCAGTAC 378
Qy 370 AGGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 429
Db 379 AGGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 438
Qy 430 AAGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 489
Db 439 AAGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 466
Qy 490 AGGCGAGCCTTGGGCTCCACACCCGCGCAAGCCAGATGTCACCTCAAGCCCGGACAAC 549
Db 467 -----CAGCCCAAGGTTGTCACCTGAGCCCGGACAAC 498

```

[illegible]

```

RESULT 15
US-60-453--050-346
: Sequence 346, Application US/60453050
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: LUKE, May
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CLO01457
: CURRENT APPLICATION NUMBER: US/60/453,050
: CURRENT FILING DATE: 2003-03-10
: NUMBER OF SEQ ID NOS: 82762
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 346
: LENGTH: 1353
: TYPE: DNA
: ORGANISM: Homo sapiens
US-60-453--050-346

```

Query Match	60.6%;	Score 952.2;	DB 11;	Length 1353;
Best Local Similarity	80.7%;	Pred. No. 1.6e-175;		
Matches 126;	Conservative	4;	Mismatches 12;	Indels 285; Gaps 2;

Qy	10	GGTGCCTTGAATGCTGTGTCGGCCCGCTCCGACCAATTTCACAACCACTNAGCAACCGGAGC	69
Db	19	GGCTTCCGATGATCTGTTCGCGCCCTCCGACCAATTTCACACACACATGACACCGGAC	78
Qy	70	ACCCAGTCTCTCTTTCTTCTGTGCTGCTCTTCACAGTCTTACATTTGTTACAGTTCT	1289
Db	79	ACCCAGTCTCTCTTCTCTGTGCTGTCTCTTCACAGTCTTACATTTGTTACAGTTCT	1389
Qy	130	GGTCATGCAAGCTCTACCCGACGATGGAGAAAGAGAACCTTCGAGTACCCAGAGATTCA	1899
Db	139	GGTCATGCAAGCTCTACCCGACGATGGAGAAAGAGAACCTTCGAGTACCCAGAGATTCA	1989
Qy	190	GGTCCACACCTCTACTAGAAAGATGCTGTAGATAGACACACAGCTACCTTCAGACAC	2499
Db	199	GTGCCAGAGCTCTACTAGAAAGATGCTGTAGATAGACACACAGCTACCTTCAGACAC	2589
Qy	250	AACCCCGGTTTCAGGCTCTCCACACATCCAGAGACAGATATGATCTGTGCGCCCGGACAG	3099
Db	259	AACCCCGGTTTCAGGCTCTCCACACATCCAGAGACAGATATGATCTGTGCGCCCGGACAG	3189
Qy	310	GAAACAAGTTCAAGTTCAAGCTTCACGACCTGAGGAGACAGATGTCACTCTGTCCAGTCAC	3699
Db	319	GAAACAAGTTCAAGTTCAAGCTTCACGACCTGAGGAGACAGATATGATCTCTGTCCAGTCAC	3789
Qy	370	AAGCAGAGCCCTGGGAGCTCCACACCCCGCCAGCCACAGATGTCACTCTAGCCCGGACAC	4299
Db	379	AAGCAGAGCCCTGGGAGCTCCACACCCCGCCAGCCACAGATGTCACTCTAGCCCGGACAC	4389
Qy	430	AAGCAGAGCCCGGGAGCTCCACCGGCGCCCGCGGCGCAAGATGTCACTCTGGCCCGGACAC	4899
Db	439	AAGCAGAGCCCGGGAGCTCCACCGGCGCCCGCGGCGCAAGATGTCACTCTGGCCCGGACAC	4989
Qy	490	AAGCAGCCCGCGGAGCTTCACACCGCCCGCGCCGCTCAAGTGTCACTCTGGCCCGGACAC	5499
Db	497	-----CAGCCCAAGATGTCACTCTGGCCCGGACAC	4989
Qy	550	AAGCAGAGCCCGGAGCTTCACACCGCCCGCGCCGCTCAAGTGTCACTCTGGCCCGGACAC	6099
Db	499	AAGCAGAGCCCGGAGCTTCACACCGCCCGCGCCGCTCAAGTGTCACTCTGGCCCGGACAC	5589
Qy	610	AAGCAGAGCTTGGAGCTTCACACCGCCCGCTCAAGTGTCAAGTGTCACTCTGGCTCAGGCTCT	6699
Db	559	AAGCAGAGCTTGGAGCTTCACACCGCCCGCTCAAGTGTCAAGTGTCACTCTGGCTCAGGCTCT	6189
Qy	670	GCATCAGGCTCAGCTTCTACTCTGTGTGCACACAGGACCTTGTGCACAGGCTACACAAAC	7299
Db	619	GCATCAGGCTCAGCTTCTACTCTGTGTGCACACAGGACCTTGTGCACAGGCTACACAAAC	6789
Qy	730	CCAGCAGACAGAGCACTTCACACAGACATTCGAGCCACACTGTGATCTCTTCAAC	7899

```

Db 679 CCAGCCAGCAAGACATCTCAATTCACATTCACAGCCACCACTGATATCTTACCCACC 738
QY 790 CTTGGCAGCCATAGACACCAAGATGATGCGATAGACATCAACATAGACAGGTACTCT 849
Db 739 CTTGCCAGCCATAGACACCAAGATGATGCGATAGACATCAACATAGACAGGTACTCT 798
QY 850 CTGACCTCTCAATCAAGACATCTTCCCAAGTGTCTACTGCGGCTCTCTTTCTTTTC 909
Db 799 CTGACCTCTCAATCAAGACATCTTCCCAAGTGTCTACTGCGGCTCTCTTTCTTTTC 858
QY 910 CTGCTTTTCAATTTCAACCTCAAGTTTAATTCCTCTGAGAAATCCAGCAGCAGAC 969
Db 859 CTGCTTTTCAATTTCAACCTCAAGTTTAATTCCTCTGAGAAATCCAGCAGCAGAC 918
QY 970 TACTACCAAGAGCTGAGAGACATTTTGAATGTTTTTGCAGATTTATTAACAAGG 1029
Db 919 TACTACCAAGAGCTGAGAGACATTTTGAATGTTTTTGCAGATTTATTAACAAGG 953
QY 1030 GGTTTCTGGGCTCTCCAAATTAAGTTCAAGGCAGAGATCTGTGGTGTACAATTGACT 1089
Db 954 ----- 953
QY 1090 CTGGCTTCAGAGAAAGTACATCAATGTCAGACGTGAGACACAGTTCAATCAGTAT 1149
Db 954 ----- 953
QY 1150 AAAACGAGAGCAGCTCTGATATTAACCTGACATCTCAGACGTGAGGTGATGATGTG 1209
Db 954 ----- 953
QY 1210 CCATTTCTTTCTGCGCCAGTCTGCGGCTGCGGCTGCGAGCTGCGGCAATGCGCTGCTG 1269
Db 954 -----GTCGCGGCTGCGGCTGCGAGCTGCGGCAATGCGCTGCTG 993
QY 1270 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1329
Db 994 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1053
QY 1330 CAGTGCCTGCGGAAAGAACTACGCGGCACTGAGCATCTTTCCAGCCCGGAGTACCTACCAT 1389
Db 1054 CAGTGCCTGCGGAAAGAACTACGCGGCACTGAGCATCTTTCCAGCCCGGAGTACCTACCAT 1113
QY 1390 CCTATGAGAGATGACCCGACCTACGACACCAATGCGGCTATGTCCTTACGAGTACC 1449
Db 1114 CCTATGAGAGATGACCCGACCTACGACACCAATGCGGCTATGTCCTTACGAGTACC 1173
QY 1450 GATGTAAGCCCTATGAGAAAGTCTTCTGAGGTAAATGTTGAGAGAGCCTCTTTACACA 1509
Db 1174 GATGTAAGCCCTATGAGAAAGTCTTCTGAGGTAAATGTTGAGAGAGCCTCTTTACACA 1233
QY 1510 AACCCAGAGTGGCAGCAGCTTCTGCAACTTGTAGGGGAGCAGTGGCCTCTGAGCTGAG 1569
Db 1234 AACCCAGAGTGGCAGCAGCTTCTGCAACTTGTAGGGGAGCAGTGGCCTCTGAGCTGAG 1293
QY 1570 TGG 1572
Db 1294 TGG 1296

```

Search completed: May 8, 2003, 18:14:50
 Job time : 797 secs


```
Db 121 ACAGGTTCTGGGTCAATGACGCTACCCAGGTGAGAAAGAGACTTCGGCTCCAG 180
Qy 181 AGAAGTTCAAGTCCCAAGCTCACTGAGAGAAATGTGTGATGTGACCAAGCACTACTC 240
Db 181 AGAAGTTCAAGTCCCAAGCTCACTGAGAGAAATGTGTGATGTGACCAAGCACTACTC 240
Qy 241 TCCAGCAACAGCCCGGTTGAGGCTCTGCAACATCGAGGGAAGAGATGCTACTGGCC 300
Db 241 TCCAGCAACAGCCCGGTTGAGGCTCTGCAACATCGAGGGAAGAGATGCTACTGGCC 300
Qy 301 CCGGCAACGGAACCAAGCTTCAAGTTCAAGTCACTGGGGAAGAGATGCTACTGGCC 360
Db 301 CCGGCAACGGAACCAAGCTTCAAGTTCAAGTCACTGGGGAAGAGATGCTACTGGCC 360
Qy 361 CCAATCAACAGCCAGGCTCGGAGCTCAACCAAGCCAGCCAGCCAGATGCTACTGAGC 420
Db 361 CCAATCAACAGCCAGGCTCGGAGCTCAACCAAGCCAGCCAGCCAGATGCTACTGAGC 420
Qy 421 CCGGCAACAGCCAGGCTCGGAGCTCAACCAAGCCAGCCAGCCAGATGCTACTGAGC 480
Db 421 CCGGCAACAGCCAGGCTCGGAGCTCAACCAAGCCAGCCAGCCAGATGCTACTGAGC 480
Qy 481 CCGGCAACAGCCAGGCTCGGAGCTCAACCAAGCCAGCCAGCCAGATGCTACTGAGC 540
Db 481 CCGGCAACAGCCAGGCTCGGAGCTCAACCAAGCCAGCCAGCCAGATGCTACTGAGC 540
Qy 514 -----CCGCGCGCCACGAGTCACTCGGCC 540
Db 541 CCGGCAACAGCCAGGCTCGGAGCTCAACCAAGCCAGCCAGCCAGATGCTACTGAGC 600
Qy 541 CCGGCAACAGCCAGGCTCGGAGCTCAACCAAGCCAGCCAGCCAGATGCTACTGAGC 600
Db 601 CCGGCAACAGCCAGGCTCGGAGCTCAACCAAGCCAGCCAGCCAGATGCTACTGAGC 660
Qy 601 CCGGCAACAGCCAGGCTCGGAGCTCAACCAAGCCAGCCAGCCAGATGCTACTGAGC 660
Db 661 CCGGCAACAGCCAGGCTCGGAGCTCAACCAAGCCAGCCAGCCAGATGCTACTGAGC 720
Qy 661 TCAAGCTCTGATCAAGGCTCAAGCTTCACTGTGATGCAACCGGCACTCTCCAGGCT 720
Db 721 TCAAGCTCTGATCAAGGCTCAAGCTTCACTGTGATGCAACCGGCACTCTCCAGGCT 780
Qy 721 ACCACAACCCAGCCAGGCTCAAGCTTCACTGTGATGCAACCGGCACTCTCCAGGCT 780
Db 781 ACCACAACCCAGCCAGGCTCAAGCTTCACTGTGATGCAACCGGCACTCTCCAGGCT 840
Qy 781 CCAACCAACCCAGCCAGGCTCAAGCTTCACTGTGATGCAACCGGCACTCTCCAGGCT 840
Db 841 CCAACCAACCCAGCCAGGCTCAAGCTTCACTGTGATGCAACCGGCACTCTCCAGGCT 900
Qy 841 GATACCTCTCTCACTCTCTCAATCAAGCACTTCCCAAGTTGCTACTGGAGCTCT 900
Db 901 GATACCTCTCTCACTCTCTCAATCAAGCACTTCCCAAGTTGCTACTGGAGCTCT 960
Qy 901 TTTCTTTCTCTCTCTCTCTCAATCAAGCACTTCCCAAGTTGCTACTGGAGCTCT 960
Db 961 TTTCTTTCTCTCTCTCTCTCAATCAAGCACTTCCCAAGTTGCTACTGGAGCTCT 1020
Qy 961 AGCAGCAGCTCTCTCTCTCTCAATCAAGCACTTCCCAAGTTGCTACTGGAGCTCT 1020
Db 1021 AGCAGCAGCTCTCTCTCTCTCAATCAAGCACTTCCCAAGTTGCTACTGGAGCTCT 1080
Qy 1021 AAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db 1081 AAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Qy 1081 CAATTGACTCTGAGCTCTCTCTCTCAATCAAGCACTTCCCAAGTTGCTACTGGAGCTCT 1140
Db 1141 CAATTGACTCTGAGCTCTCTCTCTCAATCAAGCACTTCCCAAGTTGCTACTGGAGCTCT 1200
Qy 1141 AATCAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
```

```
Db 1201 AATCAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy 1201 AGTCATGTGCAATTTCTTTCTCTGACCAAGCTGAGGAGCTGAGGAGCTGAGGAGATC 1260
Db 1261 AGTCATGTGCAATTTCTTTCTCTGACCAAGCTGAGGAGCTGAGGAGCTGAGGAGATC 1320
Qy 1261 GCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320
Db 1321 GCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
Qy 1321 GCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
Db 1381 GCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1440
Qy 1381 ACTTACCATCTATGAGGAGATACCCCACTACCAACCCATGAGGAGCTATGTGCCCT 1440
Db 1441 ACTTACCATCTATGAGGAGATACCCCACTACCAACCCATGAGGAGCTATGTGCCCT 1500
Qy 1441 AGCAGTACCGATGTGAGCCCTTATGAGAGAGTTTCTGAGAGTAAATGTGTGAGAGCTC 1500
Db 1501 AGCAGTACCGATGTGAGCCCTTATGAGAGAGTTTCTGAGAGTAAATGTGTGAGAGCTC 1560
Qy 1501 TCTTACCAAAACCCAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
Db 1561 TCTTACCAAAACCCAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
Qy 1561 TGAAGTGTAGTGG 1572
Db 1621 TGAAGTGTAGTGG 1632

RESULT 2
US-09-967-768A-224
Sequence 224: Application US/09967768A
Patent No. US2002150877A1
GENERAL INFORMATION:
APPLICANT: Augeneus, Meena
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
FILE REFERENCE: 689290-72
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: US/60/236, 109
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236, 034
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236, 111
SOFTWARE: PatentIn version 3.0
SEQ ID NO 224
LENGTH: 1721
TYPE: DNA
ORGANISM: Homo sapiens
US-09-967-768A-224

Query March 94.2%; Score 1481.2; DB 10; Length 1721;
Best Local Similarity 95.5%; Pred. No. 0;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;
```


Dh	295	GCCTCCCAACCCACGATGTCACTCGCCCTCGGACACACAGCCCGCCCTCGAGTCCAC	354
Oy	391	ACCCGCGACCCACGATATGTCACTCGCCCGCGACCAACAGCACGCCCGAGGTCAAC	450
Dh	355	GCCTCCCAACCCACGATGTCACTCGCCCTCGGACACACAGCCCGCCCTCGAGTCCAC	414
Oy	451	GCCTCCCAACCCACGATGTCACTCGCCCTCGGACACACAG-----	492
Oy	493	CCGCCCCCGGCTCCACCGCCCCCGCGCCACAGGTGTCACTCGACCCCGAACACAG	552
Dh	415	GCACCCCAACCCACGATGTCACTCGCCCTCGGACACACAGCCCGCCCTCGAGTCCAC	474
Dh	475	CCGAGCCCGGAGCTCCACCGCCCCCGAGCCACAGTGTCACTCGCCCTCGAACACAG	534
Oy	553	CCGAGCCCGGAGCTCCACCGCCCCCGAGCCACAGTGTCACTCGCCCTCGAACACAG	612
Dh	535	CCGAGCCCGGAGCTCCACCGCCCCCGAGCCACAGTGTCACTCGCCCTCGAACACAG	594
Oy	613	CCGAGCTTGGGCTCCACCGCCCTCCAGTGTCACTCGACCTCGAGCTGTGCA	672
Dh	595	CCGAGCTTGGGCTCCACCGCCCTCCAGTGTCACTCGACCTCGAGCTGTGCA	654
Oy	673	TCAGGCTAGCTTCTACTGTGGTCAACAACGGACCTGTGCAAGGGCTTACCAACCCCA	732
Dh	655	TCAGGCTAGCTTCTACTGTGGTCAACAACGGACCTGTGCAAGGGCTTACCAACCCCA	714
Oy	733	GCACGACAGAGCACTCCACCCGAGATCCCGACGACACTCTGATATCTCTACACCCCT	792
Dh	715	GCACGACAGAGCACTCCATCTCAATTCGACGACCACTCTGATATCTCTACACCCCT	774
Oy	793	GCACGACATAGACAAAGCTGATGAGCGACATGACATCCACATAGACAGGTACCTCTC	852
Dh	775	GCACGACATAGACAAAGCTGATGAGCGATGACATGACATGACAGGTACCTCTCTC	834
Oy	853	ACCTCTTCATCAACGACACTCTCCCAATGTCTACAGGGCTCTCTCTCTCTCTCTG	912
Dh	835	ACCTCTTCATCAACGACACTCTCCCAATGTGTCTAGGGGCTCTCTCTCTCTCTCTG	894
Oy	913	TCCTTTCACATTTCAAACTCCAGTTTAAATTCCTCTGGAAGATCCACGACCGACTAC	972
Dh	895	TCCTTTCACATTTCAAACTCCAGTTTCTCTCTCTCTCGAAGATCCACGACCGACTAC	954
Oy	973	TACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTCGAGATTATTAACAAGGGGT	1032
Dh	955	TACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTCGAGATTATTAACAAGGGGT	1014
Oy	1033	TTTTCGGAGCTCTCCAAATTAAAGTTCAGGCGCAGATCTGTGTGGTAAATTAAGACTGT	1092
Dh	1015	TTTTCGGAGCTCTCTCCAAATTAAAGTTCAGGCGCAGATCTGTGTGGTAAATTAAGACTGT	1074
Oy	1093	GCCTTCGAGAAGGTACATCAATGTCAAGGTGAGACAGAGTTCAATCAATATATAA	1154
Dh	1075	GCCTTCGAGAAGGTACATCAATGTCAAGGTGAGACAGAGTTCAATCAATATATAA	1134
Oy	1153	AGGGAAGAGGCTCTCGATATTAACCTGACAGCTCCAGAGCTGACAGTACATGATCA	1214
Dh	1135	AGGGAAGAGGCTCTCGATATTAACCTGACAGCTCCAGAGCTGACAGTACATGATCA	1194
Oy	1213	TTTTTCCTTCTGTCCAGAGCTCGGGGCTCGGGGTCCAGAGCTGAGGGGCTACCGGTCTGGTG	1272
Dh	1195	TTTTTCCTTCTGTCCAGAGCTCGGGGCTCGGGGTCCAGAGCTGAGGGGCTACCGGTCTGTG	1254
Oy	1273	CTGAGTCTGTCTCTGAGTGTGCGTGTGCAATGTGTATCTCAATGCTGTGCTGTGCA	1334
Dh	1255	CTGAGTCTGTCTCTGAGTGTGCGTGTGCAATGTGTATCTCAATGCTGTGCTGTGCA	1314
Oy	1333	TGCGCGCAAAAGACATACGGGACGTGTGACATTTTCCAGCCGGGATATCTACATCTCT	1394
Dh	1315	TGCGCGCAAAAGACATACGGGACGTGTGACATTTTCCAGCCGGGATATCTACATCTCT	1374
Oy	1393	ATGAGGAGTACCCCACTTACCAACCCATGGGCGCTATGTGCCCTTAGACATACCAT	1455

Accession	Sequence	Position
Dd	ATGAGCGAGTACCCCACTACACACCATGGGCGCTATATCCCCCTTAGAGTACCGAT	1432
Qy	CGTAGCCCTTATGAAAGTTCTTCGAGGTTATGTTGTGACGACGCTCTTTTACAACAC	1512
Dd	CGTACCCCTTATGAAAGTTCTTCGAGGTTATGTTGTGACGACGCTCTTTTACAACAC	1494
Qy	CGAGTAGGCGACCACTTTCGCACTTGAG	1545
Dd	CGAGTAGGCGACCACTTTCGCACTTGAG	1527

RESULT 5
US-09-96

```

? Sequence 105:Application US/09964824A
? Patent No. US20020102531A1
? GENERAL INFORMATION:
? APPLICANT: Horrigan, Stephen
? TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
? TITLE OF INVENTION: Sess
? FILE REFERENCE: 689290-73
? CURRENT APPLICATION NUMBER: US/09/964,824A
? CURRENT FILING DATE: 2001-09-27
? PRIOR APPLICATION NUMBER: US/60/236,033
? PRIOR FILING DATE: 2000-09-28
? PRIOR APPLICATION NUMBER: US/60/236,032
? PRIOR FILING DATE: 2000-09-28
? PRIOR APPLICATION NUMBER: US/60/236,028
? PRIOR FILING DATE: 2000-09-28
? NUMBER OF SEQ ID NOS: 563
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 105
? LENGTH: 4139
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-964-824A-105

```

Query Match	75.7%;	Score 1190.2;	DB 10;	Length 4139;
Best Local Similarity	94.8%;	Pred. No. 0;		
Matches 1231; Conservative	0;	Mismatches 68;	Indels 0;	Gaps 0;

QY	309	GGAAACAAGCTTTCAGGTTTCAGCTGTCACATCTGGGAGACAGAGATGTCACTCTGGATCCCAATGCAC	368
Db	2545	CAGGCGCGGCGCCCGGGGCTCCACCGGCCCCCAGCCAGCATGATGTCACTCTGGGCGCCCGGAGCAC	2600
QY	369	CAGGCGAAGCCTTGGGGTCCACACCGCGCAGCGCCAGATGATCACTTCAGGCGCCGAGACAA	428
Db	2665	CAGGCGGCGCTCCGGGCTCCACCGGCGCCCGCCAGCTCCAGATGTCACTCTGGGCGCCGAGCAC	2722
QY	429	CAGGCGAAGCCTCCGGGCTCCACCGGCGCCCGCGCCAGATGTCACTCTGGGCGCCGAGCAC	488
Db	2725	CAGGCGGCGCTCCGGGCTCCACCGGCGCCCGCCAGCTCCAGATGTCACTCTGGGCGCCGAGCAC	2788
QY	489	CAGGCGCGCCCGCGGGTCCACCGGCGCCCGCGCCAGATGTCACTCTGGGCGCCGAGCAC	548
Db	2785	CAGGCGGCGCTCCGGGCTCCACCGGCGCCCGCCAGCTCCAGATGTCACTCTGGGCGCCGAGCAC	2844
QY	549	CAGGCGGCGCTCCGGGCTCCACCGGCGCCCGCGCCAGATGTCACTCTGGGCGCCGAGCAC	608
Db	2845	CAGGCGGCGCTCCGGGCTCCACCGGCGCCCGCCAGCTCCAGATGTCACTCTGGGCGCCGAGCAC	2900
QY	609	CAGGCGGCGCTTGGGGTCCACCGGCGCCCGCTCCAGATGTCACTCTGGGCTCCAGGCTC	668
Db	2905	CAGGCGCGCTTGGGGTCCACCGGCGCCCGCTCCAGATGTCACTCTGGGCTCCAGGCTC	2966
QY	669	TGCATTCAGGCTCAGGCTTTCATCTGTGTGCAACAAGGAGCTCTGCGCAGGGGCTACCAAC	728
Db	2965	TGCATTCAGGCTCAGGCTTTCATCTGTGTGCAACAAGGAGCTCTGCGCAGGGGCTACCAAC	3022
QY	729	CCGAGCCAGCAAGAGCATTCACCCAGACATTCACAGCCACACATCTGTGATATCTCTACAC	788

Db 3025 CCCAGGACAGAGAGCACTCATTCCTCAATTCACGACCACTCATTCATCTTACCTTACAC 3084
Qy 789 CTTTGCACAGCATTAGACACCAAGACTGATGACATGACATGACATGACATGACATGAC 848
Db 3085 CTTTGCACAGCATTAGACACCAAGACTGATGACATGACATGACATGACATGACATGAC 3144
Qy 849 TTTCACTCTCTCAATCAGACATCTTCCGAGTGTCTACGAGGAGTCTCTTCTTTT 908
Db 3145 TTTCACTCTCTCAATCAGACATCTTCCGAGTGTCTACGAGGAGTCTCTTCTTTT 3204
Qy 909 CTTTGCATTTTCAATTTCAATCTCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 968
Db 3205 CTTTGCATTTTCAATTTCAATCTCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 3264
Qy 969 CTTTGCATTTTCAATTTCAATCTCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1028
Db 3265 CTTTGCATTTTCAATTTCAATCTCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 3324
Qy 1029 GGGTTTCTGGGCTCTTCAATTTAAGTTCAGGACAGATCTGTGTGTACATTTGAC 1088
Db 3325 GGGTTTCTGGGCTCTTCAATTTAAGTTCAGGACAGATCTGTGTGTACATTTGAC 3384
Qy 1089 TTTGACCTTCCGAGAGGTACATCATATGTCAGAGATGAGACACAGTTCAATCAGTA 1148
Db 3385 TTTGACCTTCCGAGAGGTACATCATATGTCAGAGATGAGACACAGTTCAATCAGTA 3444
Qy 1149 TAAAAAGAGAGAGCTCTGATATACCTGACATCTGAGAGTCAAGCTGACGATCAGT 1208
Db 3445 TAAAAAGAGAGAGCTCTGATATACCTGACATCTGAGAGTCAAGCTGACGATCAGT 3504
Qy 1209 GCGATTTCTTCT 1268
Db 3505 GCGATTTCTTCT 3564
Qy 1269 GGT 1328
Db 3565 GGT 3624
Qy 1329 TCAGTCCGCGAAAGAACTACGAGGAGCTGAGACATCTTTCACGCGCGAGTACATCA 1388
Db 3625 TCAGTCCGCGAAAGAACTACGAGGAGCTGAGACATCTTTCACGCGCGAGTACATCA 3684
Qy 1389 TCTTATGACGAGTACCCACCTACACACCATGAGGCTATGTGTGTGTGTGTGTGTGT 1448
Db 3685 TCTTATGACGAGTACCCACCTACACACCATGAGGCTATGTGTGTGTGTGTGTGTGT 3744
Qy 1449 CGATGTAGCCCTTATGAGAGGTTTCTGCAAGTAACTGTGTGTGTGTGTGTGTGT 1508
Db 3745 CGATGTAGCCCTTATGAGAGGTTTCTGCAAGTAACTGTGTGTGTGTGTGTGTGT 3804
Qy 1509 AAACCCAGATGCGACCCCTTCTCCCACTTTGTAAGG 1547
Db 3805 AAACCCAGATGCGACCCCTTCTCCCACTTTGTAAGG 3843

RESULT 6
US-09-964-824A-578
Sequence 578, Application US/09964824A
Patent No. US20020102531A1
GENERAL INFORMATION:
APPLICANT: Horizon, Stephen
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
FILE REFERENCE: 689290-73
CURRENT APPLICATION NUMBER: US/09/964, 824A
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: US/60/236, 033
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236, 032
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236, 028
PRIOR FILING DATE: 2000-09-28

NUMBER OF SEQ ID NOS: 583
SOFTWARE: PatentIn version 3.0
SEQ ID NO 578
LENGTH: 4139
TYPE: DNA
ORGANISM: Homo sapiens
US-09-964-824A-578
Query: Match 75.7%; Score 1190.2; DB 10; Length 4139;
Best Local Similarity 94.8%; Pred. No. 0; Mismatches 68; Indels 0; Gaps 0;
Matches 1331; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 249 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 308
Db 2545 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 3264
Qy 309 GGAACCAAGCTTCAAGTTTCACTGCTCACTGAGGACAGATGTCACTGTGCTCCAGTAC 368
Db 2605 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 2664
Qy 369 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 428
Db 2665 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 2724
Qy 429 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 488
Db 2725 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 2784
Qy 489 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 548
Db 2785 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 2844
Qy 549 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 608
Db 2845 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 2904
Qy 609 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 668
Db 2905 CAGCCCGGTTTCAAGGCTCTCTCAACACTCAGAGACAGATGTCATCTGTGCCCCGAC 2964
Qy 669 TGCATCAGGCTCAGCTTCT 728
Db 2965 TGCATCAGGCTCAGCTTCT 3024
Qy 729 CCGAGCCGAGAGAGACCTCAGCAGATTCAGAGCCAGCAGCTGTGATCTCTTACAC 788
Db 3025 CCGAGCCGAGAGAGACCTCAGCAGATTCAGAGCCAGCAGCTGTGATCTCTTACAC 3084
Qy 789 CTTTGCACAGCATTAGACACCAAGACTGATGACATGACATGACATGACATGACATGAC 848
Db 3085 CTTTGCACAGCATTAGACACCAAGACTGATGACATGACATGACATGACATGACATGAC 3144
Qy 849 TTTCACTCTCTCAATCAGACATCTTCCGAGTGTCTACGAGGAGTCTCTTCTTTT 908
Db 3145 TTTCACTCTCTCAATCAGACATCTTCCGAGTGTCTACGAGGAGTCTCTTCTTTT 3204
Qy 909 CTTTGCATTTTCAATTTCAATCTCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 968
Db 3205 CTTTGCATTTTCAATTTCAATCTCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 3264
Qy 969 CTTTGCATTTTCAATTTCAATCTCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1028
Db 3265 CTTTGCATTTTCAATTTCAATCTCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 3324
Qy 1029 GGGTTTCTGGGCTCTTCAATTTAAGTTCAGGACAGATCTGTGTGTACATTTGAC 1088
Db 3325 GGGTTTCTGGGCTCTTCAATTTAAGTTCAGGACAGATCTGTGTGTACATTTGAC 3384
Qy 1089 TTTGACCTTCCGAGAGGTACATCATATGTCAGAGATGAGACACAGTTCAATCAGTA 1148
Db 3385 TTTGACCTTCCGAGAGGTACATCATATGTCAGAGATGAGACACAGTTCAATCAGTA 3444
Qy 1149 TAAAAAGAGAGAGCTCTGATATACCTGACATCTGAGAGTCAAGCTGACGATCAGT 1208

Db 3445 TAAACGGAAGCAGCTCTCATATACTGACATCTGACAGCTGACCGTAGTAGT 3504
 Qy 1209 GCCATTCTCTTCTGTGCGCAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGT 1268
 Db 3505 GCCATTCTCTTCTGTGCGCAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGT 3564
 Qy 1269 GGTGCTGT 1328
 Db 3565 GGTGCTGT 3624
 Qy 1329 TCAGTCCGCGAAGAACTACGAGGAGCTGACATCTTCAGCGCGAGATCCATACA 1388
 Db 3625 TCAGTCCGCGAAGAACTACGAGGAGCTGACATCTTCAGCGCGAGATCCATACA 3684
 Qy 1389 TCCTATGACGAGTACCCCACTACACACCCATGAGGCTATGTGCCCTAGCAGTAC 1448
 Db 3685 TCCTATGACGAGTACCCCACTACACACCCATGAGGCTATGTGCCCTAGCAGTAC 3744
 Qy 1449 CGATCGTAGCCCTATGAGAGTTCGTGAGGTATGAGTGGACAGAGCTCTTACAC 1508
 Db 3745 CGATCGTAGCCCTATGAGAGTTCGTGAGGTATGAGTGGACAGAGCTCTTACAC 3804
 Qy 1509 AAACCCAGCAGTGGCAGCAGCTTCTGCCAATTGTAGG 1547
 Db 3805 AAACCCAGCAGTGGCAGCAGCTTCTGCCAATTGTAGG 3843

RESULT 7
 US-09-864-864-334
 ; Sequence 334, Application US/09864864
 ; Patent No. US20020102679A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Dillon, Devin C.
 ; APPLICANT: Secrist, Heather
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Fling, Steve P.
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Benson, Darin R.
 ; APPLICANT: Carter, Derrick
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
 ; FILE REFERENCE: 210121.523
 ; CURRENT APPLICATION NUMBER: US/09/864,864
 ; NUMBER OF SEQ ID NOS: 341
 ; SOFTWARE: Corixa Invention Disclosure Database
 ; SEQ ID NO 334
 ; LENGTH: 4139
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-864-864-334

Query Match 75.7%; Score 1190.2; DB 10; Length 4139;
 Best Local Similarity 94.8%; Pred. No. 0; Mismatches 68; Indels 0; Gaps 0;
 Matches 1211; Conservative 0;

Qy 249 CAGCCCGGTTGAGGCTCTTCACCACTGAGGAGCAGATGTCACTTGGCCCGGAC 308
 Db 2545 CAGGCGGCGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2604
 Qy 309 GGAACGAGTTGAGGTTGAGTGCACCTGAGGAGCAGATGTCACTTGGCCCGGAC 368
 Db 2605 CAGGCGGCGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2664
 Qy 369 CAGGCGGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 428
 Db 2665 CAGGCGGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2724

Qy 429 CAGGCGGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 488
 Db 2725 CAGGCGGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2784
 Qy 489 CAGGCGGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 548
 Db 2785 CAGGCGGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2844
 Qy 549 CAGGCGGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 608
 Db 2845 CAGGCGGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2904
 Qy 609 CAGGCGGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 668
 Db 2905 CAGGCGGCGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2964
 Qy 669 TGCAATCAGGCTCAGCTTCTACTCTGTGTGCAACCGGCACTCTGCGCGGCTACCAAC 728
 Db 2965 TGCAATCAGGCTCAGCTTCTACTCTGTGTGCAACCGGCACTCTGCGCGGCTACCAAC 3024
 Qy 729 CCAGCGGCGAAGAGCTCCAGCCGAGCTTCCAGCCAGCACTGTATCTCTACAC 788
 Db 3025 CCAGCGGCGAAGAGCTCCAGCTTCCAGCCAGCACTGTATCTCTACAC 3084
 Qy 789 CCTTGCGAGCATGAGCAAGATGATGCGAGTGCATCAGCATGAGCAGTACTCC 848
 Db 3085 CCTTGCGAGCATGAGCAAGATGATGCGAGTGCATCAGCATGAGCAGTACTCC 3144
 Qy 849 TCTCACTCTCTCCATCAGAGCACTTCTCCAGTGTCTACTGAGGCTCTTTT 908
 Db 3145 TCTCACTCTCTCCATCAGAGCACTTCTCCAGTGTCTACTGAGGCTCTTTT 3204
 Qy 909 CCTGCTTTTCACTTCAAACTCAGTTAATTTCTCTGGAAGATCCAGACGGA 968
 Db 3205 CCTGCTTTTCACTTCAAACTCAGTTAATTTCTCTGGAAGATCCAGACGGA 3264
 Qy 969 CTACTCAAGAGCTGAGAGAGACATTTGGAATGTTTTCAGATTATTAACAAG 1028
 Db 3265 CTACTCAAGAGCTGAGAGAGACATTTGGAATGTTTTCAGATTATTAACAAG 3324
 Qy 1029 GGGTTTTCTGGGCTCTCCAAATATTAAGTCAAGGAGATCTGAGTGTCAATGAC 1088
 Db 3325 GGGTTTTCTGGGCTCTCCAAATATTAAGTCAAGGAGATCTGAGTGTCAATGAC 3384
 Qy 1089 TCTGCGCTTCCGAGAGGTAACATGATGTCCAGAGCTGAGACACAGTTCAATGATA 1148
 Db 3385 TCTGCGCTTCCGAGAGGTAACATGATGTCCAGAGCTGAGACACAGTTCAATGATA 3444
 Qy 1149 TAAACGAGAGAGCTCTCGATATTAACCTGAGATCTGAGAGTGAAGTCAATGT 1208
 Db 3445 TAAACGAGAGAGCTCTCGATATTAACCTGAGATCTGAGAGTGAAGTCAATGT 3504
 Qy 1209 GCATTTCTTTCTCTGCGCAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGT 1268
 Db 3505 GCATTTCTTTCTCTGCGCAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGT 3564
 Qy 1269 GGTGCTGT 1328
 Db 3565 GGTGCTGT 3624
 Qy 1329 TCAATGCGCGAAGAACTACGAGGAGCTGACATCTTCAGCGCGAGATCCATACA 1388
 Db 3625 TCAATGCGCGCGAAGAACTACGAGGAGCTGACATCTTCAGCGCGAGATCCATACA 3684
 Qy 1389 TCCTATGACGAGTACCCCACTACACACCCATGAGGCTATGTGCCCTAGCAGTAC 1448
 Db 3685 TCCTATGACGAGTACCCCACTACACACCCATGAGGCTATGTGCCCTAGCAGTAC 3744
 Qy 1449 CGATCGTAGCCCTATGAGAGTTCGTGAGGTATGAGTGGACAGAGCTCTTACAC 1508
 Db 3745 CGATCGTAGCCCTATGAGAGTTCGTGAGGTATGAGTGGACAGAGCTCTTACAC 3804
 Qy 1509 AAACCCAGCAGTGGCAGCAGCTTCTGCCAATTGTAGG 1547

Db 3805 AACCCAGCAGTGGCAGCCGCTTCTGCCAATTGTAGGG 3843

```

RESULT 8
US-09-880-107-2121
Sequence 2121, Application US/0980107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2121
LENGTH: 4139
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 J05582
US-09-880-107-2121

```

Query Match	75.7%;	Score 1190.2;	DB 10;	Length 4139;
Best Local Similarity	94.8%;	Pred. No. 0;		
Matches 1231; Conservative	0;	Mismatches 68;	Indels 0;	Gaps 0

Qy	249	CAGCCCCGTTGAGGCTCTTCAACAATCAGAGGACAGATGTCACTCTGGACCCCGGACAC	308
Db	2445	CAGGCGGAGCCCGGACTCAACCGCCCCCGACGCCAGGTGTCACTTGGACCCGGACAC	2604
Qy	309	GGAACCAAGCTTCAGGTTCAAGCTGTCACTTGGGAGCAGATGTCACTTCGGTTCACATTCAC	368
Db	2605	CAGGCGGAGCCCGGACTCAACCGCCCCCGACGCCAGGTGTCACTTGGACCCGGACAC	2664
Qy	369	CAGGCGAGCCCTGGGCTTCAACCAACCCCGGAGCCACATGTCACTTACGCCCGGACAA	428
Db	2665	CAGGCGGAGCCCGGACTCAACCGCCCCCGACGCCAGGTGTCACTTGGACCCCGGACAC	2724
Qy	429	CAAGCAGAGCCCGGACTCAACCGCCCCCGGAGCCACGGTGTCACTTGGACCCCGGACAC	488
Db	2725	CAGGCGGAGCCCGGACTCAACCGCCCCCGACGCCAGGTGTCACTTGGACCCCGGACAC	2788
Qy	489	CAGGCGGAGCCCGGACTCAACCGCCCCCGGAGCCACGGTGTCACTTGGACCCCGGACAC	548
Db	2785	CAGGCGGAGCCCGGACTCAACCGCCCCCGACGCCAGGTGTCACTTGGACCCCGGACAC	2844
Qy	549	CAGGCGGAGCCCGGACTCAACCGCCCCCGGAGCCACGGTGTCACTTGGACCCCGGACAA	608
Db	2845	CAGGCGGAGCCCGGACTCAACCGCCCCCGACGCCAGGTGTCACTTGGACCCCGGACAA	2904
Qy	609	CAGGCGGAGCCCTTGGACTCAACCGCCCCCTTCAGTTCACATGTCACTTGGACTCAGAGCTC	668
Db	2905	CAGGCGGAGCCCTTGGACTCAACCGCCCCCTTCAGTTCACATGTCACTTGGACTCAGAGCTC	2964
Qy	669	TGCATCAGGCTCAGCTTTATCTCTGTGTGCACAACGGACCTTGTGCAGGGTTCACAAC	728
Db	2965	TGCATCAGGCTCAGCTTTATCTCTGTGTGCACAACGGACCTTGTGCAGGGTTCACAAC	3024
Qy	729	CCGAGCAGCAGAGACCTCACCCACAGATTTCCAGGCCACCACTCTGATTTCTCTACAC	788
Db	3025	CCGAGCAGCAGAGACCTCATTTCTAATTTCCAGCCACCACTCTGATTTCTCTACAC	3088
Qy	789	CTTGTGCAGCCATAGACAACAAGCTGATGTGCAGTAGACTCAACCATAGACAGGTACTTCC	848

Db	3085	CCTTGGCAGCCATTAACACCAAGACTATATGCAATGACATCAACATATAGCTCGGTACCTCC	3144
Qy	849	TTCTACCTTCCTTCATTAACAAGCACTTCTCCCAAGTTGTCTATCTGAGGCTCTCTTCTCTTTT	908
Db	3145	TTCTACCTCTCTCCAAATACAGAGCACTTCTCCCAAGTTGTCTATCTGAGGCTCTCTTCTCTTTT	3204
Qy	909	CCGTGCTTTTCACATTTCAAAACCTCCAGTTTAATCTCTCTGGAAGATCCAGACACGA	968
Db	3205	CCGTGCTTTTCACATTTCAAAACCTCCAGTTTAATCTCTCTGGAAGATCCAGACACGA	3264
Qy	969	CTATCTACCAAGACTGCAAGAGACATTTCTGAAATGTTTTTGCAGATTTATTAACAAAG	1028
Db	3265	CTATCTACCAAGACTGCAAGAGACATTTCTGAAATGTTTTTGCAGATTTATTAACAAAG	3324
Qy	1029	GAGTTTTTCAGGGCCCTCTCAATATTTAAGTTCAAGGCCAGATCTGTGTGTCAATTAAGC	1088
Db	3325	GAGTTTTTCAGGGCCCTCTCAATATTTAAGTTCAAGGCCAGATCTGTGTGTCAATTAAGC	3384
Qy	1089	TCGTGGCTTCGGAAGAGTACATCAATGTCCAGACGTGGAGACACAGTTCATTAACATGA	1148
Db	3385	TCGTGGCTTCGGAAGAGTACATCAATGTCCAGACGTGGAGACACAGTTCATTAACATGA	3444
Qy	1149	TAAACGGAAGACAGCTCTCGATATATACTGACATCGATCGACAGCTGAGTCAATGT	1208
Db	3445	TAAACGGAAGACAGCTCTCGATATATACTGACATCGATCGACAGCTGAGTCAATGT	3504
Qy	1209	GCCATTTCTCTTCTCTCTGCCCAGTCTGGGGCTGGGGGTGCAGAGCTGGGGACATGGCGTCT	1268
Db	3505	GCCATTTCTCTTCTCTCTGCCCAGTCTGGGGCTGGGGGTGCAGAGCTGGGGACATGGCGTCT	3564
Qy	1269	GGTGCTGTCTGT	1328
Db	3565	GGTGCTGTCTGT	3624
Qy	1329	TCAAGTCCGCGCGAAGAAATCTAAGAGGAGTGAACATCTTTCAGCCCGGGATATACCTACA	1388
Db	3625	TCAAGTCCGCGCGAAGAAATCTAAGAGGAGTGAACATCTTTCAGCCCGGGATATACCTACA	3684
Qy	1389	TCCTATGAGCAGTACCCCACTTACACACCCATGGGCGTATGTGCCCCCTTAGCAGTAC	1448
Db	3685	TCCTATGAGCAGTACCCCACTTACACACCCATGGGCGTATGTGCCCCCTTAGCAGTAC	3744
Qy	1449	CGATGTGAGCCCTCTATAGAAAGTTTCTGCAAGTATGATGTGGACAGAGCTCTCTTACAC	1508
Db	3745	CGATGTGAGCCCTCTATAGAAAGTTTCTGCAAGTATGATGTGGACAGAGCTCTCTTACAC	3804
Qy	1509	AAACCCAGCAGTGGAGCCACTTCTGCCAATCTGTAGGG	1547
Db	3805	AAACCCAGCAGTGGAGCCACTTCTGCCAATCTGTAGGG	3843
RESULT 9			
US-09-925-301-19			
Sequence 19, Application US/09925301			
Patent No. US20020052308A1			
GENERAL INFORMATION:			
APPLICANT: Rosen et al.			
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies			
FILE REFERENCE: PA106			
CURRENT APPLICATION NUMBER: US/09/925,301			
PRIOR APPLICATION NUMBER: PCT/US00/05862			
PRIOR FILING DATE: 2000-03-08			
PRIOR APPLICATION NUMBER: 60/124,270			
NUMBER OF SEQ ID NOS: 1999-03-12			
NUMBER OF SEQ ID NOS: 1694			
SOFTWARE: PatentIn Ver. 2.0			
SEQ ID NO 19			
LENGTH: 1355			
TYPE: DNA			
ORGANISM: Homo sapiens			
FEATURE:			
NAME/KEY: misc feature			

Query Match 20.9%; Score 328.4; DB 9; Length 330;

Best Local Similarity 99.7%; Fred. No. 9.9e-90;

Matches 329; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 968 ACTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTTGGAGATTTATTAACAAG 1027

DB 1 ACTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTTGGAGATTTATTAACAAG 60

QY 1028 GGGGTTTTCTGGGCTCTCCATATTTAAGTTCAGGCCAGATCTGGTGGTGAATTTGA 1087

DB 61 GGGGTTTTCTGGGCTCTCCATATTTAAGTTCAGGCCAGATCTGGTGGTGAATTTGA 120

QY 1088 CTCTGGCTTCCGAGAGGATACATCATATTCACGACGTGAGACACAGTTCAATCAT 1147

DB 121 CTCTGGCTTCCGAGAGGATACATCATATTCACGACGTGAGACACAGTTCAATCAT 180

QY 1148 ATMAAACGAGAGAGCTCTGCATATTAACCTGACATCTCAGACGTGAGATG 1207

DB 181 ATMAAACGAGAGAGCTCTGCATATTAACCTGACATCTCAGACGTGAGATG 240

QY 1208 TGCATTTCTTTCTGCCCCAGTCTGGGGCTGGGGGCGAGGCTGGGGCATCGGCTGC 1267

DB 241 TGCATTTCTTTCTGCCCCAGTCTGGGGCTGGGGGCGAGGCTGGGGCATCGGCTGC 300

QY 1268 TGGTCTGCTGTGTGTCTGTTGTTGGCTGG 1297

DB 301 TGGTCTGCTGTGTGTCTGTTGTTGGCTGG 330

RESULT 14

US-10-017-754-9

; Sequence 9, Application US/10017754

; Publication No. US20030054363A1

; GENERAL INFORMATION:

; APPLICANT: Henderson, Robert A.

; APPLICANT: Wang, Tongcong

; APPLICANT: Watanabe, Yoshihiro

; APPLICANT: Johnson, Jeffrey C.

; APPLICANT: Retter, Marc W.

; APPLICANT: Marnetakis, Margarita

; APPLICANT: Carter, Derrick

; APPLICANT: Fanger, Gary R.

; APPLICANT: Vedavick, Thomas S.

; APPLICANT: McNabb, Andria

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.478C18

; CURRENT APPLICATION NUMBER: US/10/017.754

; CURRENT FILING DATE: 2001-10-29

; NUMBER OF SEQ ID NOS: 2004

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 9

; LENGTH: 330

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-017-754-9

Query Match 20.9%; Score 328.4; DB 9; Length 330;

Best Local Similarity 99.7%; Fred. No. 9.9e-90;

Matches 329; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 968 ACTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTTGGAGATTTATTAACAAG 1027

DB 1 ACTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTTGGAGATTTATTAACAAG 60

QY 1028 GGGGTTTTCTGGGCTCTCCATATTTAAGTTCAGGCCAGATCTGGTGGTGAATTTGA 1087

DB 61 GGGGTTTTCTGGGCTCTCCATATTTAAGTTCAGGCCAGATCTGGTGGTGAATTTGA 120

QY 1088 CTCTGGCTTCCGAGAGGATACATCATATTCACGACGTGAGACACAGTTCAATCAT 1147

DB 121 CTCTGGCTTCCGAGAGGATACATCATATTCACGACGTGAGACACAGTTCAATCAT 180

QY 1148 ATMAAACGAGAGAGCTCTGCATATTAACCTGACATCTCAGACGTGAGATG 1207

DB 181 ATMAAACGAGAGAGCTCTGCATATTAACCTGACATCTCAGACGTGAGATG 240

QY 1208 TGCATTTCTTTCTGCCCCAGTCTGGGGCTGGGGGCGAGGCTGGGGCATCGGCTGC 1267

DB 241 TGCATTTCTTTCTGCCCCAGTCTGGGGCTGGGGGCGAGGCTGGGGCATCGGCTGC 300

QY 1268 TGGTCTGCTGTGTGTCTGTTGTTGGCTGG 1297

DB 301 TGGTCTGCTGTGTGTCTGTTGTTGGCTGG 330

RESULT 15

US-09-918-995-29996/c

; Sequence 29996, Application US/09918995

; Publication No. US20030073623A1

; GENERAL INFORMATION:

; APPLICANT: HySeq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995

; CURRENT FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US/09/235,076

; NUMBER OF SEQ ID NOS: 38054

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 29996

; LENGTH: 364

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-918-995-29996

Query Match 18.7%; Score 294.2; DB 9; Length 364;

Best Local Similarity 97.4%; Fred. No. 2.6e-79;

Matches 299; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 112 ACAAGTTTACAGAGTTCTGATCATGCAAGCTTACCCCGAGTGGAGAAAGAGACTTGG 171

DB 307 ACAAGTTTACAGAGTTCTGATCATGCAAGCTTACCCCGAGTGGAGAAAGAGACTTGG 248

QY 172 GCTACCCAGAGAGTTCAAGTCCAGGCTCTACTGAGAGAAAGTGTGATGACCAAG 231

DB 247 GCTACCCAGAGAGTTCAAGTCCAGGCTCTACTGAGAGAAAGTGTGATGACCAAG 188

QY 232 AGCTTATCTCCAGCCAGCCAGCCGCTTCAAGGCTCTTCAACTCAGGAGCAGATGTC 291

DB 187 AGCTTATCTCCAGCCAGCCAGCCGCTTCAAGGCTCTTCAACTCAGGAGCAGATGTC 128

QY 292 ACTTGAGCCCGGCGACGAGAAACGAGTTTCAAGTTTCAAGTCCAGCTGGGAGCAGATGTC 351

DB 127 ACTTGAGCCCGGCGACGAGAAACGAGTTTCAAGTTTCAAGTCCAGCTGGGAGCAGATGTC 68

QY 352 ACCCTGGTCCAGTCAAGGAGCAGGCTGGGCTCCAGCCCGGAGCCAGATGTC 411

DB 67 ACCCTGGTCCAGTCAAGGAGCAGGCTGGGCTCCAGCCCGGAGCCAGATGTC 8

QY 412 ACTTCAG 418

DB 7 ACTTCAG 1

Search completed: May 8, 2003, 18:19:15

Job time : 281 secs

Thu May, 8 18:53:09 2003

us-09-658-621b-1.inp

Page 1

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 15:04:32 ; Search time 3749 Seconds
(without alignments)
10542.542 Million cell updates/sec

Title: US-09-658-621B-1
Perfect score: 1572
Sequence: 1 gaattccctgcgtcgtgaa.....tcgacctcgcagctgagtcgg 1572
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues
Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents NA Main:*

1: /cgn2_6/ptodata/1/pna/US0955A_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US0956A_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US0957A_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US0958A_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US0959A_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US0960A_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US0961A_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US0962A_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US0963A_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US0964A_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US0965A_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US0966A_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US0967A_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US0968A_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US0969A_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US0970A_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US0971A_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US0972A_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US0973A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US0974A_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US0975A_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US0976A_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US0977A_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US0978A_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US0979A_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US0980A_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US0981A_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US0982A_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US0983A_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US0984A_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US0985A_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US0986A_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US0987A_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US0988A_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US0989A_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US0990A_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US0991A_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US0992A_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US0993A_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US0994A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US0995A_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US0996A_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US0997A_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1572	100.0	25	US-09-658-621-1
2	1481.8	94.3	41	US-10-198-846-1589
3	1481.2	94.2	1	PCT-US02-07826-211
4	1481.2	94.2	1	PCT-US02-18947-775
5	1481.2	94.2	33	US-09-864-864-280
6	1481.2	94.2	36	US-09-867-768A-224
7	1481.2	94.2	38	US-10-007-926A-58
8	1481.2	94.2	1721	US-10-029-517-3
9	1481.2	94.2	1721	US-10-097-340-211
10	1481.2	94.2	1721	US-10-171-311-155
11	1481.2	94.2	1721	US-10-172-118-775
12	1481.2	94.2	2678	US-10-452-157-103
13	1481.2	94.2	62	US-60-209-009-103
14	1481.2	94.2	28	US-09-716-473-2290
15	1454.4	92.5	29	US-09-726-172-2117
16	1454.4	89.8	38	US-09-864-824A-573
17	1412.2	89.8	1804	US-10-029-517-17
18	1412.2	89.8	1823	PCT-US02-08456-339
19	1342	85.4	1428	US-08-479-537-1
20	1309.8	83.3	8	US-08-479-537-1
21	1289.2	82.0	8	US-08-479-537-2

```
22 1257 80.0 1527 17 US-09-366-670-19 Sequence 19, Appl
23 1257 80.0 1527 39 US-10-057-136-19 Sequence 19, Appl
24 1213.4 77.2 1737 32 US-09-830-513-11 Sequence 11, Appl
25 1210.2 77.0 1371 32 US-09-830-513-5 Sequence 5, Appl
26 1190.2 75.7 4139 1 PCT-US02-19669-310 Sequence 310, Appl
27 1190.2 75.7 4139 33 US-09-864-864-334 Sequence 211, Ap
28 1190.2 75.7 4139 36 US-09-880-107-2121 Sequence 105, App
29 1190.2 75.7 4139 36 US-09-964-824A-105 Sequence 157, App
30 1190.2 75.7 4139 41 US-10-171-311-157 Sequence 310, App
31 1190.2 75.7 4139 41 US-10-177-293-310 Sequence 5270, Ap
32 1180.2 75.6 4139 65 US-60-213-360-5270 Sequence 2741, Ap
33 1186.8 75.5 4168 17 US-09-205-070-2741 Sequence 2741, Ap
34 1186.8 75.5 4168 17 US-09-340-623-2741 Sequence 2741, Ap
35 1186.8 75.5 4168 17 US-09-898-888-2741 Sequence 2741, Ap
36 1186.8 75.5 4168 33 US-09-898-888-2741 Sequence 2741, Ap
37 1186.8 75.5 4168 33 US-09-898-888-2741 Sequence 2741, Ap
38 1043.6 66.4 2297 32 US-09-856-988A-41 Sequence 19, Appl
39 978 62.2 1355 1 PCT-US00-05882-19 Sequence 19, Appl
40 978 62.2 1355 34 US-09-925-301-19 Sequence 7468, Ap
41 946.2 60.2 1261 28 US-09-705-256A-7468 Sequence 10940, A
42 946.2 60.2 1261 60 US-60-164-285-7468 Sequence 1, Appl
43 919.4 58.5 3361 71 US-60-278-258-10840 Sequence 1, Appl
44 750 47.7 6286 1 PCT-US01-30151-1 Sequence 19, Appl
45 750 47.7 8186 38 US-10-029-517-19
```

ALIGNMENTS

```
RESULT 1
US-09-658-621-1
; Sequence 1, Application US/09658621
; GENERAL INFORMATION:
; APPLICANT: Taylor-Papadimitriou, Joyce
; APPLICANT: Henkamp, Lukas Carl
; APPLICANT: Offitger, Rieck
; APPLICANT: Melief, Cornelis Johanna Maria
; APPLICANT: Acres, Bruce
; APPLICANT: Thomas, Mireille
; TITLE OF INVENTION: MUC-1 derived peptides
; FILE REFERENCE: 029395-017
; CURRENT APPLICATION NUMBER: US/09/658, 621
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/187,215
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: GB 9921242.5
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: EP 99 40 2237.4
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patencin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)..(1542)
US-09-658-621-1
```

Query Match 100.0%; Score 1572; DB 25; Length 1572;

Best Local Similarity 100.0%; Pred. No. 8.9e-296; Matches 1572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAATTCCTGCTGCTGTAATCTGTTGCCCCCTCCCAACCCATTTCACACACCATG 60
Db 1 GAATTCCTGCTGCTGTAATCTGTTGCCCCCTCCCAACCCATTTCACACACCATG 60
QY 61 AACACGGGACACCAAGTCTCTTCTCTGCTGCTGCTCTCAAGTGTGTTT 120
Db 61 AACACGGGACACCAAGTCTCTTCTCTGCTGCTGCTCTCAAGTGTGTTT 120
QY 121 ACAGGTTCTGCTGCAAGCTCTACCCAGGTGAGAAAGAGACTTCGGCTACCCAG 180
```

```
|||||
Db 121 ACAGGTTCTGCTGCAAGCTCTACCCAGGTGAGAAAGAGACTTCGGCTACCCAG 180
QY 181 AGAAGTTCAAGTCCAGCTCTCACTGAGAAATGCTGAGTATGACACAGCTACTC 240
Db 181 AGAAGTTCAAGTCCAGCTCTCACTGAGAAATGCTGAGTATGACACAGCTACTC 240
QY 241 TCAGAGCAAGCCCGGTTGAGCTCTCACTCAACACTGAGGAGACAGATGCTGAGCC 300
Db 241 TCAGAGCAAGCCCGGTTGAGCTCTCACTCAACACTGAGGAGACAGATGCTGAGCC 300
QY 301 CCGGCAAGGAAACAGCTTCAAGTTAGTCTGACCTGAGGAGACAGATGCTGAGCTC 360
Db 301 CCGGCAAGGAAACAGCTTCAAGTTAGTCTGACCTGAGGAGACAGATGCTGAGCTC 360
QY 361 CCAATCAACAGGACAGCCCTGAGCTCTCAACCCCGGACCCAGATGCTCAAGCC 420
Db 361 CCAATCAACAGGACAGCCCTGAGCTCTCAACCCCGGACCCAGATGCTCAAGCC 420
QY 421 CCGGCAAGAACAGCTCCGAGCTCTCAACCCCGGACCCAGATGCTCAAGCTGAGCC 480
Db 421 CCGGCAAGAACAGCTCCGAGCTCTCAACCCCGGACCCAGATGCTCAAGCTGAGCC 480
QY 481 CCGGCAAGAACAGCTCCGAGCTCTCAACCCCGGACCCAGATGCTCAAGCTGAGCC 540
Db 481 CCGGCAAGAACAGCTCCGAGCTCTCAACCCCGGACCCAGATGCTCAAGCTGAGCC 540
QY 541 CCGGCAAGAACAGCTCCGAGCTCTCAACCCCGGACCCAGATGCTCAAGCTGAGCC 600
Db 541 CCGGCAAGAACAGCTCCGAGCTCTCAACCCCGGACCCAGATGCTCAAGCTGAGCC 600
QY 601 CCGGCAAGAACAGCTCCGAGCTCTCAACCCCGGACCCAGATGCTCAAGCTGAGCC 660
Db 601 CCGGCAAGAACAGCTCCGAGCTCTCAACCCCGGACCCAGATGCTCAAGCTGAGCC 660
QY 661 TCAGGCTCTGATCAAGCTCAAGCTTCTAATCTGAGTGCACAAAGGACCTCTGAGGCT 720
Db 661 TCAGGCTCTGATCAAGCTCAAGCTTCTAATCTGAGTGCACAAAGGACCTCTGAGGCT 720
QY 721 ACCAACAACCCAGCCAGCAAGACACTTCCACCCAGATTCACAGCACTCTGATACT 780
Db 721 ACCAACAACCCAGCCAGCAAGACACTTCCACCCAGATTCACAGCACTCTGATACT 780
QY 781 CCTACACCTTGGCAGGACATPACACAAAGCTAGTGCAGTGCAGCTCAACATGACAG 840
Db 781 CCTACACCTTGGCAGGACATPACACAAAGCTAGTGCAGTGCAGCTCAACATGACAG 840
QY 841 GTCCTCTCTCACTCTCTGATGACAGACTTCTCCAGTGTCTAATCTCTGAGATCC 900
Db 841 GTCCTCTCTCACTCTCTGATGACAGACTTCTCCAGTGTCTAATCTCTGAGATCC 900
QY 901 TTTCTTTTCTGCTCTTTCATCAATTCAACTCCAGTTTAACTCTCTGAGATCC 960
Db 901 TTTCTTTTCTGCTCTTTCATCAATTCAACTCCAGTTTAACTCTCTGAGATCC 960
QY 961 AGCAGCGACTACTACCAAGAGCTGAGAGAGACTTCTGAAATGTTTGGAGATTAT 1020
Db 961 AGCAGCGACTACTACCAAGAGCTGAGAGAGACTTCTGAAATGTTTGGAGATTAT 1020
QY 1021 AAACAGAGGGGTTTCTGAGGCTCTGCAATATTAAGTTCAAGGACAGATCTGAGTGA 1080
Db 1021 AAACAGAGGGGTTTCTGAGGCTCTGCAATATTAAGTTCAAGGACAGATCTGAGTGA 1080
QY 1081 CAATGACTCTGAGCTTTCAGAGAGGTACATCAATGTCAGAGCTGAGAGACAGTTC 1140
Db 1081 CAATGACTCTGAGCTTTCAGAGAGGTACATCAATGTCAGAGCTGAGAGACAGTTC 1140
QY 1141 AATCAGTATTAACAGGAAGAGCTCTGATATTAACCTGACGATCTCAAGCTGAGCTG 1200
Db 1141 AATCAGTATTAACAGGAAGAGCTCTGATATTAACCTGACGATCTCAAGCTGAGCTG 1200
QY 1201 AGTCATGAGCAATTCCTTCTCTGCTGAGCTGAGGCTGAGGCTGAGGCTGAGCTG 1260
Db 1201 AGTCATGAGCAATTCCTTCTCTGCTGAGCTGAGGCTGAGGCTGAGGCTGAGCTG 1260
```

```
Db 1201 AGTCATGTGCAATTCCTTTCTCTGCGCACTGCGGGGCTGGGGGCTGCCAGGCTGGGCAATC 1260
Oy 1261 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db 1261 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Oy 1321 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Db 1321 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Oy 1381 ACCTACCATTCCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 ACCTACCATTCCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Oy 1441 AGCAGTACCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 1500
Db 1441 AGCAGTACCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 1500
Oy 1501 TCTTACCAAAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Db 1501 TCTTACCAAAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Oy 1561 TGAGCTGAGTGG 1572
Db 1561 TGAGCTGAGTGG 1572

RESULT 2
US-10-198-846-12589
; Sequence 12589, Application US/10198846
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRL-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12589
; LENGTH: 2026
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1, 2026
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-12589

Query Match 94.3%, Score 1491.8; DB 41; Length 2026;
Best Local Similarity 95.6%; Pred. No. 4.3e-280;
Matches 1559; Conservative 0; Mismatches 12; Indels 60; Gaps 1;

Oy 2 AATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 61
Db 10 AATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 69
Oy 62 CACCGGAGACCCAGTCTCTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 121
Db 70 CACCGGAGACCCAGTCTCTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 129
Oy 122 CAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181
Db 130 CAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 189
Oy 182 GAAGTTGAGTGGCAGCTCTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
Db 182 GAAGTTGAGTGGCAGCTCTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241

Db 190 GAAGTTGAGTGGCAGCTCTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 249
Oy 242 CAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
Db 250 CAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
Oy 302 CAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
Db 310 CAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 369
Oy 362 CAGTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
Db 370 CAGTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
Oy 422 CAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 481
Db 430 CAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 489
Oy 482 CAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 513
Db 490 CAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 519
Oy 514 -----CCGCGCGCCAGGAGTCACTCGGCCC 541
Db 550 CAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 609
Oy 542 CAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 601
Db 610 CAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 669
Oy 602 CAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661
Db 670 CAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729
Oy 662 CAGGCTCTGATGAGGCTGAGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 721
Db 730 CAGGCTCTGATGAGGCTGAGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 789
Oy 722 CACAAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 781
Db 790 CACAAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 849
Oy 782 CTACCAACCTTGGCAGCATGAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841
Db 850 CTACCAACCTTGGCAGCATGAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 909
Oy 842 TACCTCTGATGAGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 901
Db 910 TACCTCTGATGAGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 969
Oy 902 TCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 961
Db 970 TCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1029
Oy 962 GCACCGATCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021
Db 1030 GCACCGATCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1089
Oy 1022 AACCAAGGAGGTTTCTGAGGCTCTCCAAATTTAAGTTCAAGGAGAGAGAGAGAGAGAGAGAG 1081
Db 1090 AACCAAGGAGGTTTCTGAGGCTCTCCAAATTTAAGTTCAAGGAGAGAGAGAGAGAGAGAGAG 1149
Oy 1082 AATGACCTGAGGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
Db 1150 AATGACCTGAGGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1209
Oy 1142 ATCAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
Db 1210 ATCAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1269
Oy 1202 GTCATGTGCAATTTCTTTCTGCGCAGTCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCT 1261
Db 1270 GTCATGTGCAATTTCTTTCTGCGCAGTCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCT 1329
```


Qy	1201	AGCAATGGGCATTTCTCTTCTCTGCGCAGTCGCGGAGCTGGGGATGCGAGGCTGGGGGATC	1260
Db	1261	AGTAGATGGCATTTCTCTTCTCTGCGCAATCTGGGGCTGGGGATGCGAGGCTGGGGGATC	1320
Qy	1261	GCGCTGCTGGTCTGCTGCTCTGTGTTCTGTGTGGCTGGCGCATGTGTATCTCATTTGGCTTG	1320
Db	1321	GCGCTGCGGTGCTGCTGCTGTGTGTTCTGTGTGGCTGGCGCATGTGTATCTCATTTGGCTTG	1380
Qy	1321	GCTGTCTGTCAAGTGGCGCGCAAGAAATCAAGGAGCAGCTGAGCATCTTTCTCAGCCCGGAT	1380
Db	1381	GCTGTCTGTCAAGTGGCGCGCAAGAAATCAAGGAGCAGCTGAGCATCTTTCTCAGCCCGGAT	1440
Qy	1381	ACCTACATCTTATGAGAGGATGACCCACCTACACACCTATGGAGCTATGATGACCCCTT	1440
Db	1441	ACCTACATCTTATGAGAGGATGACCCACCTACACACCTATGGAGCTATGATGACCCCTT	1500
Qy	1441	AGCAGTACCGATCGTAGCCCTTATGAGAAAGTTTCTGCAAGGTATGTGTGGACGACGCTC	1500
Db	1501	AGGAGTACCGATCGTAGCCCTTATGAGAAAGTTTCTGCAAGGTATGTGTGGACGACGCTC	1560
Qy	1501	TCTTACACAAACCCAGACAGTGGCAGCACTTGTGCAACTTGTAGAGGAGAGTGGCCCTC	1560
Db	1561	TCTTACACAAACCCAGAGATGGCAGCACTTGTGCAACTTGTAGAGGAGAGTGGCCCTC	1620
Qy	1561	TGAGCTGAGTGG 1572	
Db	1621	TGAGCTGAGTGG 1632	

```

RESULT 4
PCT-US02-18947-775
Sequence 775, Application PC/TUS0218947
GENERAL INFORMATION:
APPLICANT: Rosettta Impharmatics
TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
FILE REFERENCE: 9301-175-228
CURRENT APPLICATION NUMBER: PCT/US02/18947
CURRENT FILING DATE: 2002-06-14
PRIOR APPLICATION NUMBER: 60/380,770
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ. ID NOS: 2659
SEQ. ID NO 775
LENGTH: 1721
TYPE: DNA
ORGANISM: Homo sapiens
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: NM_002456
DATABASE ENTRY DATE: 2001-06-18
PCT-US02-18947-775

```

Query Match	94.2%;	Score 1481.2;	DB 1;	Length 1721;
Best Local Similarity	95.5%;	Pred. No. 5.5e-280;		
Matches 1559; Conservative	0;	Mismatches 13;	Indels 60;	Gaps 1;

QY	1	GAATTCCTGGAGTCTGAAATCTGTTTGGCCCCCTCCACCACTTTCAACACACATG	60
Db	1	GAATTCCTGGAGTCTGAAATCTGTTTGGCCCCCTCCACCACTTTCAACACACATG	60
QY	61	ACACGGGAGACCCAGTCTCTTTTCTTCTGTGTGTCTCTCACAGTGTACAGTGT	120
Db	61	ACACGGGAGACCCAGTCTCTTTTCTTCTGTGTGTCTCTCACAGTGTACAGTGT	120
QY	121	ACAGGTTCTGTGTATGTACAGTCTTACCCAGTGTGAGAAAGAGACTTGTGGTACCG	180
Db	121	ACAGGTTCTGTGTATGTACAGTCTTACCCAGTGTGAGAAAGAGACTTGTGGTACCG	180
QY	181	AGAGATTCAGTGCCCAAGCTCTTACTGTAGAAGAAATGCTGTAGATATGACACAGCTTATC	240
Db	181	AGAGATTCAGTGCCCAAGCTCTTACTGTAGAAGAAATGCTGTAGATATGACACAGCTTATC	240
QY	241	TCCAGCCACAGTCCCGGTTACAGCTTCTTCACACTTACAGGACAGAGATGACCTTGGCC	300
Db	241	TCCAGCCACAGTCCCGGTTACAGCTTCTTCACACTTACAGGACAGAGATGACCTTGGCC	300

[illegible]

Qy	1321	GGCTGCTCATGAGCGCCGCCAAAAGA	CTACGGGACAGCTGACATCTTTTTCAGAGCCCGGAA	1380
Db	1381	GCTGCTCATGAGCGCCGCCAAAAGA	CTACGGGACAGCTGACATCTTTTTCAGAGCCCGGAA	1440
Qy	1381	ACCTACATCCCTTATGAGGAGATACCCAC	CTCACACCATATGGGCGCTATGAGCCCTC	1440
Db	1441	ACCTACATCCCTTATGAGGAGATACCCAC	CTCACACCATATGGGCGCTATGAGCCCTC	1500
Qy	1441	AGCAGTACCAATGATGCCCCCTTATGAA	AAATTTTCTGAGATTAATCTGAGGAGAGAC	1500
Db	1501	AGCAGTACCAATGATGCCCCCTTATGAA	AAATTTTCTGAGATTAATCTGAGGAGAGAC	1560
Qy	1501	TCTTTCACAAACCCACACATGACGAC	CACTTCTGCAACTTTATGAGGAGACGTCGCC	1560
Db	1561	TCTTTCACAAACCCACACATGACGAC	CACTTCTGCAACTTTATGAGGAGACGTCGCC	1620
Qy	1561	TGAGCTGAGTGG	1572	
Db	1621	TGAGCTGAGTGG	1632	

```

RESULT 5
US-09-864-864-280
Sequence 280, Application US/09864864 .
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Dillon, David C.
APPLICANT: Secrist, Heather
APPLICANT: Lodes, Michael J.
APPLICANT: Algate, Paul A.
APPLICANT: Fling, Steve P.
APPLICANT: Mannion, Jane P.
APPLICANT: Benson, Darin R.
APPLICANT: Carter, Derrick
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.523
CURRENT APPLICATION NUMBER: US/09/864, 864
CURRENT FILING DATE: 2001-05-23
NUMBER OF SEQ ID NOS: 341
SOFTWARE: Cortix Invention Disclosure Database
SEQ ID NO 280
LENGTH: 1721
TYPE: DNA
ORGANISM: Homo sapiens
US-09-864-864-280

```

Query Match	Similarity	94.2%	Score	1481.2	DB 33	Length	1721
Best Local	Similarity	95.5%	Pred.	Low	5.5e-28		
Matches	1559	Conservative	0	Mismatches	13	Indels	60
							Gaps
							1
Qy	1	GAATTCCCTGGCTGCTGAAATCTGTTCTGCCCCCTCCCAACCCATTCCACCAACCATG	60				
Db	1	GAATTCCTCCGCTCTTGAATCTGTTCTGCCCCCTCCCAACCCATTCCACCAACCATG	60				
Qy	61	ACACCGGGACACCCAGTCTCTCTTCTTCTGCTGCTGCTCCACAGTGTTCACAGTTGTT	120				
Db	61	ACACCGGGACACCCAGTCTCTCTTCTTCTGCTGCTGCTCCACAGTGTTCACAGTTGTT	120				
Qy	121	ACAGGTTCTGGTCATGCACAGCTCTACCCAGGTGGAGAAAGAGACTTCGAGTACCAG	180				
Db	121	ACAGGTTCTGGTCATGCACAGCTCTACCCAGGTGGAGAAAGAGACTTCGAGTACCAG	180				
Qy	181	AGAAGTTCAAGTACCACACTCTACGTGAGAGAAATGCTGTGATGTGACAGCAGGTACTC	240				
Db	181	AGAAGTTCAAGTACCACACTCTACGTGAGAGAAATGCTGTGATGTGACAGCAGGTACTC	240				
Qy	241	TCCAGCACAAGCCCGGTTTCAAGCTCTTCCACACTCAGGGACAGAGTCTCACTTGTGCC	300				
Db	241	TCCAGCACAAGCCCGGTTTCAAGCTCTTCCACACTCAGGGACAGAGTCTCACTTGTGCC	300				

[illegible]

Thu May, 8 18:53:09 2003

us-09-658-621b-1.rnp

Page 7

Db	1381	GCTGTCGTCAATGCGCCGCGAAGAACTACGAGGAGCTGGACATCTTTCACACCCGGAGAT	1440
Qy	1381	ACCTACATCCATATGAGCGAGATACCCACCTTACGACACCCATGGGAGCTATATGCCCCCT	1440
Db	1441	ACCTTACATCTCTATGAGGAGTATACCCACCTTACACACCCATGGAGGCTATATGCCCCCT	1500
Qy	1441	AGCAGTACCGATCGTAGGCCCTTATGAAAGTTTCTGCAGGTAATGATGCGACGACCTTC	1500
Db	1501	AGCAGTACCGATCGTAGGCCCTTATGAAAGTTTCTGCAGGTAATGATGCGACGACCTTC	1560
Qy	1501	TCTTACACAAACCCAGCAGGTGCGACCATTTGTGCACTTGTATGAGGAGCACGTGCCCCCTC	1560
Db	1561	TCTTACACAAACCCAGCAGGTGCGACCATTTGTGCACTTGTATGAGGAGCACGTGCCCCCTC	1620
Qy	1561	TGAGCTGAGTGG	1572
Db	1621	TGAGCTGAGTGG	1632

```

RESULT 6
US-09-967-768A-224
? Sequence 224, Application US/03967768A
? GENERAL INFORMATION:
? APPLICANT: Augustus, Meena
? TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signature
? TITLE OF INVENTION: Seqs
? FILE REFERENCE: 689290-72
? CURRENT APPLICATION NUMBER: US/09/967,768A
? CURRENT FILING DATE: 2001-09-28
? PRIOR APPLICATION NUMBER: US/60/236,109
? PRIOR FILING DATE: 2000-09-28
? PRIOR APPLICATION NUMBER: US/60/236,034
? PRIOR FILING DATE: 2000-09-28
? PRIOR APPLICATION NUMBER: US/60/236,111
? PRIOR FILING DATE: 2000-09-28
? NUMBER OF SEQ ID NOS: 325
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 224
? LENGTH: 1721
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-967-768A-224

```

Query Match	94.2%	Score 181.2	DB 36	Length 1721
Best Local Similarity	95.5%	Pred. No. 5-55-280		
Matches 1559	Conservative 0	Mismatches 13	Indels 60	Gaps 1
QY 1	GAATTCCTGGCTGCTTGAATCTGTCTGCCCCCTCCACCCATTTCACACACACATG	60		
Db 1	GAATTCCTGGCTGCTTGAATCTGTCTGCCCCCTCCACCCATTTCACACACACATG	60		
QY 61	ACACCGGAGCACCACTCTCTTCTCTGCTGCTGCTCTCAAGTGCCTACAGTGT	120		
Db 61	ACACCGGAGCACCACTCTCTTCTCTGCTGCTGCTCTCAAGTGCCTACAGTGT	120		
QY 121	ACAGGTTTCAGTATGCAAGCTCTACCCAGAGTGAGAAAGAGAGATTCGGTACCAG	180		
Db 121	ACAGGTTTCAGTATGCAAGCTCTACCCAGAGTGAGAAAGAGAGATTCGGTACCAG	180		
QY 181	AGAAGTTCAGTGCCTCAAGCTCTAAGAGAAATGCTGAGATATGACACAGACGCTACTC	240		
Db 181	AGAAGTTCAGTGCCTCAAGCTCTAAGAGAAATGCTGAGATATGACACAGACGCTACTC	240		
QY 241	TTCAGCCACAGGCCCGGTTTCAGGCTCTCCACCACTCAAGGAGACAGATGTCATCTGGCC	300		
Db 241	TTCAGCCACAGGCCCGGTTTCAGGCTCTCCACCACTCAAGGAGACAGATGTCATCTGGCC	300		
QY 301	CCGGCCACCGGAACAGACTTCAGGTTTCAGGCTGCACTGGGGACAGATATGTCACCTCGGTC	360		
Db 301	CCGGCCACCGGAACAGACTTCAGGTTTCAGGCTGCACTGGGGACAGATATGTCACCTCGGTC	360		
QY 361	CGAATCACAGGCGCAGCCCTGGGCTTCACCAACCCCGGACAGCCACGATGTCACTTAGCC	420		

Dh	361	CCAGTGTCCAGAGGCGAGGCTTGTGGGCTCCACACACCCCGGAGGCCACGATGTCACTTCAGGC	420
Qy	421	CCGGACAAACAGAGCGAGCCCGGGGCTCCACCCGCCCCCGGAGCCACGATGTCACTTCAGGC	480
Dh	421	CCGGACAAACAGAGCGAGCCCGGGGCTCCACCCGCCCCCGAGGCCACGATGTCACTTCAGGC	480
Qy	481	CCGAGACACAGAGCGGCGCCCGGGGCTCCAGCGGC-----513	513
Dh	481	CCGAGACACAGAGCGGCGCCCGGGGCTCCAGCGGC-----513	513
Qy	514	-----CCGGCGGCCACGATGTCACTTCAGGC	540
Dh	541	CCGGACACACAGAGCGGCGCCCGGGGCTCCACCGGCGCCGACGCCACAGATGTCACTTCAGGC	600
Qy	541	CCGGACACACAGAGCGGCGCCCGGGGCTCCACCGGCGCCCGGAGCCACGATGTCACTTCAGGC	600
Dh	601	CCGGACACACAGAGCGGCGCCCGGGGCTCCACCGGCGCCCGAGCCACGATGTCACTTCAGGC	660
Qy	601	CCGGACAAACAGAGCGGCGCCCGGGGCTCCACCGGCGCCCTCCAGTCCACAGATGTCACTTCAGGC	660
Dh	661	CCGGACAAACAGAGCGGCGCCCGGGGCTCCACCGGCGCCCTCCAGTCCACAGATGTCACTTCAGGC	720
Qy	661	TGAGGCTCTGCATCAGGCTCAGCTTACTCTGGTGGACACAGGACCTCTCCAGAGGCT	720
Dh	721	TGAGGCTCTGCATCAGGCTCAGCTTACTCTGGTGGACACAGGACCTCTCTCCAGAGGCT	780
Qy	721	ACGACAAACCCGAGCGAGACGACATCCACCCAGCATTCGAGCCACACTCTGACT	780
Dh	781	ACGACAAACCCGAGCGAGACGACATCCATCTTCATTCGAGCCACACTCTGACT	840
Qy	781	CCCTACACCCCTTTCGAGCGACATAGACACAGACTAGTTCGAGTACCTCACATACAG	840
Dh	841	CCCTACACCCCTTTCGAGCGACATAGACACAGACTAGTTCGAGTACCTCACATACAG	900
Qy	841	GTACCTCTCTCAGCTCTCTCCAGATACAGACATTCCTCCAGATGTCTACCTGGGGTCTCT	900
Dh	901	GTACCTCTCTCAGCTCTCTCCAGATACAGACATTCCTCCAGATGTCTACCTGGGGTCTCT	960
Qy	901	TTCTTTTCTCTGTCTTTTCAATTCGAACTCCAGTTTAACTCTCTCTGGAAATCCC	960
Dh	961	TTCTTTTCTCTGTCTTTTCAATTCGAACTCCAGTTTAACTCTCTCTGGAAATCCC	1020
Qy	961	AGCACGACACTACCAAGAGCTTCAGAGACATTTCTGAATGTTTTTCAGATTAAT	1020
Dh	1021	AGCACGACACTACCAAGAGCTTCAGAGACATTTCTGAATGTTTTTCAGATTAAT	1080
Qy	1021	AAACAGAGGGGTTTTCTGGGCTCTCCAAATATAAGTTCCAGGCGAGATCTGTGGTGA	1080
Dh	1081	AAACAGAGGGGTTTTCTGGGCTCTCCAAATATAAGTTCCAGGCGAGATCTGTGGTGA	1140
Qy	1081	CAATGACCTGGGCTTCGAGAGAGGTACCAATATGTCACGAGCTGGAGACACAGTTC	1140
Dh	1141	CAATGACCTGGGCTTCGAGAGAGGTACCAATATGTCACGAGCTGGAGACACAGTTC	1200
Qy	1141	AATCAGTATTAACCGAGAGCGGCTCTGCATATTAACCTTGACATCTCAGAGTCAGGCTC	1200
Dh	1201	AATCAGTATTAACCGAGAGCGGCTCTGCATATTAACCTTGACATCTCAGAGTCAGGCTC	1260
Qy	1201	AAGTCATGTCATTTCTCTTCTCTGGCCAGTCTGGGGCTGGGGCTGGAGCTGGGGCATC	1260
Dh	1261	AAGTCATGTCATTTCTCTTCTCTGGCCAGTCTGGGGCTGGGGCTGGAGCTGGGGCATC	1320
Qy	1261	GGCGTGCAGGATGTCTGGTCTGTGTCTGTGGTGGCGGCGCAATGTCTCATTCAGGCTTG	1380
Dh	1321	GGCGTGCAGGATGTCTGGTCTGTGTCTGTGGTGGCGGCGCAATGTCTCATTCAGGCTTG	1380
Qy	1321	GGTGTGTGTAGAGCGGCGCGAAGAACTAGAGGCGAGTCACTTTCAGGCGGGGAT	1380
Dh	1381	GGTGTGTGTAGAGCGGCGCGAAGAACTAGAGGCGAGTCACTTTCAGGCGGGGAT	1440
Qy	1381	ACCTACACCTTATGAGGAGTACCCACCTTACACACCATGGGCTATGSCCCT	1440

Dh	361	CCAGTGTCCAGAGGCGAGGCTTGTGGGCTCCACACACCCCGGAGGCCACGATGTCACTTCAGGC	420
Qy	421	CCGGACAAACAGAGCAGGCCCCGGGGCTCCACCGGCCCCCGGGCCACAGTGTCACTTCGAGCC	480
Dh	421	CCGGACAAACAGAGCAGGCCCCGGGGCTCCACCGGCCCCCGGGCCACAGTGTCACTTCGAGCC	480
Qy	481	CCGAGACACAGGCGGCCCCGGGGCTCCACCGGC-----513	513
Dh	481	CCGAGACACAGGCGGCCCCGGGGCTCCACCGGC-----513	513
Qy	514	-----CCGGCGGCCACGATGTCACTTCGAGCC	540
Dh	541	CCGGACACACAGGCGGCCCCGGGGCTCCACCGGCCCCCGAGCCACAGGATGTCACTTCGAGCC	600
Qy	541	CCGGACACACAGGCGGCCCCGGGGCTCCACCGGCCCCCGAGCCACAGGATGTCACTTCGAGCC	600
Dh	601	CCGGACACACAGGCGGCCCCGGGGCTCCACCGGCCCCCGAGCCACAGGATGTCACTTCGAGCC	660
Qy	601	CCGGACACACAGGCGGCCCCGGGGCTCCACCGGCCCCCGAGCCACAGGATGTCACTTCGAGCC	660
Dh	661	CCGGACACACAGGCGGCCCCGGGGCTCCACCGGCCCCCGAGCCACAGGATGTCACTTCGAGCC	720
Qy	661	CCGGACACACAGGCGGCCCCGGGGCTCCACCGGCCCCCGAGCCACAGGATGTCACTTCGAGCC	720
Dh	721	TCAGGCTCTGCATCAGGCTCAGCTTCTACTCTGGTGGACACAGGACCTCTCGCAGAGGCT	780
Qy	721	TCAGGCTCTGCATCAGGCTCAGCTTCTACTCTGGTGGACACAGGACCTCTCGCAGAGGCT	780
Dh	781	ACCCACACCCCGAGCCGAGACGACCTCCACCCAGCATTCCTCCAGGCGACACCTCTGATTACT	840
Qy	781	ACCCACACCCCGAGCCGAGACGACCTCCACCCAGCATTCCTCCAGGCGACACCTCTGATTACT	840
Dh	841	CCCTACACCCCTTTCGAGCAGCATAGACACAGACTAGTATCCGATGAGCTACCATTAACAG	900
Qy	841	CCCTACACCCCTTTCGAGCAGCATAGACACAGACTAGTATCCGATGAGCTACCATTAACAG	900
Dh	901	GTACCTCTCTCAGCTCTCTCCAAATACAGACATTCCTCCAGATGTCTACCTGGGGTCTCT	960
Qy	901	GTACCTCTCTCAGCTCTCTCCAAATACAGACATTCCTCCAGATGTCTACCTGGGGTCTCT	960
Dh	961	TTCTTTTTCCTGTCTTTTTCAGATTTCAAACTCCAGTTTAATTCCTCTCTGGAAATCCC	1020
Qy	961	TTCTTTTTCCTGTCTTTTTCAGATTTCAAACTCCAGTTTAATTCCTCTCTGGAAATCCC	1020
Dh	1021	AGCACACGACTACTACCAAGAGCTTCGAGAGACATTTCTGAATGTTTTTTCAGATTAAT	1080
Qy	1021	AGCACACGACTACTACCAAGAGCTTCGAGAGACATTTCTGAATGTTTTTTCAGATTAAT	1080
Dh	1081	AAACAGGGGGTTTTTCTGGGCTCTCCAAATATAAGTTTCAGGCGAGAGATCTGTGGTGGTA	1140
Qy	1081	AAACAGGGGGTTTTTCTGGGCTCTCCAAATATAAGTTTCAGGCGAGAGATCTGTGGTGGTA	1140
Dh	1141	CAATGACTCTGGCCTTCGAGAGAGGTACCAATATCTCCAGATGTGGAGACACAGTTTC	1200
Qy	1141	CAATGACTCTGGCCTTCGAGAGAGGTACCAATATCTCCAGATGTGGAGACACAGTTTC	1200
Dh	1201	AATTCAGTATATAACGAGAGAGGCTCTCGATATTAACCTTGACATCTCAGAGTCAAGCTG	1260
Qy	1201	AATTCAGTATATAACGAGAGAGGCTCTCGATATTAACCTTGACATCTCAGAGTCAAGCTG	1260
Dh	1261	AGTATATGACATTTCTCTTCTCTGCGCAGGTCTGGGGCTGGGGGTGCGAGGCTGGGGCATC	1320
Qy	1261	AGTATATGACATTTCTCTTCTCTGCGCAGGTCTGGGGCTGGGGGTGCGAGGCTGGGGCATC	1320
Dh	1321	GGCGCTGCGGT	1380
Qy	1321	GGCGCTGCGGT	1380
Dh	1381	GGT	1440
Qy	1381	GGT	1440
Dh	1441	ACCTTACATCTTATGAGGAGTACCCACCTTACACACCATGTGGCTATGTGGCCCT	1500
Qy	1441	ACCTTACATCTTATGAGGAGTACCCACCTTACACACCATGTGGCTATGTGGCCCT	1500

Accession	Sequence	Length
Db	1441 ACCTAACCATCTCTATGAGGAGATACCCCACTTACACACCATGGGGGTATGATGCCCTT	1500
Qy	1441 ACAGATACCATCTATGAGCCCTTATGAGAAAGTTTCTGAGATTAATGATGACAGACCTC	1500
Db	1501 ACAGATACCATCTATGAGCCCTTATGAGAAAGTTTCTGAGATTAATGATGACAGACCTC	1560
Qy	1501 TCTTACACAAACCCAGACATGTGGACCACTTGTGCACAACTTGAAGGGACACATGTGCCCTC	1560
Db	1561 TCTTACACAAACCCAGACATGTGGACCACTTGTGCACAACTTGAAGGGACATGTGCCCTC	1620
Qy	1561 TGAGCTGATGGG 1572	
Db	1621 TGAGCTGATGGG 1632	

```

RESULT 7
US-10-007-926A-58
? Sequence 58, Application US/10007926A
? GENERAL INFORMATION:
? APPLICANT: BERTUCCI, FRANCOIS
? APPLICANT: HODIGATTE, REMI
? APPLICANT: BIRNBAUM, DANIEL
? APPLICANT: NGUYEN, CATHERINE
? APPLICANT: VIENS, PATRICE
? APPLICANT: FERT, VINCENT
? TITLE OF INVENTION: GENE EXPRESSION PROFITING OF PRIMARY BREAST CARCINOMAS
? TITLE OF INVENTION: USING ARRAYS OF CANDIDATE GENES
? FILE REFERENCE: 1546-R-00
? CURRENT APPLICATION NUMBER: US/10/007,926A
? CURRENT FILING DATE: 2001-12-07
? PRIOR APPLICATION NUMBER: 60/254,090
? PRIOR FILING DATE: 2000-12-08
? NUMBER OF SEQ ID NOS: 468
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 58
? LENGTH: 1721
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: mucin 1, transmembrane (MUC1) gene.
US-10-007-926A-58

```

Query Match	94.2%; Score 1481.2; DB 38; Length 1721;
Best Local Similarity	95.5%; Seed No. 5.5e-280;
Matches 1559; Conservative	0; Mismatches 13; Indels 60; Gaps 1
QY 1 GAATTCCTGGCGCTGTAATCTGTTGCCCCCTCCGCCACCATTTACACACACACATG 60	
Db 1 GAATTCCTGGCTGCTTAATCTGTGTCCCCCTCCGCCACCATTTACACACACACATG 60	
QY 61 ACACCGGAGCACCCAGTCTCTCTTCTCTGCTGCTCTCTCAAGTCTTACAGTTGTT 120	
Db 61 ACACCGGAGCACCCAGTCTCTCTTCTCTGCTGCTCTCTCAAGTCTTACAGTTGTT 120	
QY 121 ACAGGTTCTGTGTCATGACAGCTCACCCAGGTGAGAAAGAGACATCTGGCTACCG 180	
Db 121 ACAGGTTCTGTGTCATGACAGCTCACCCAGGTGAGAAAGAGACATCTGGCTACCG 180	
QY 181 AGAAGTTCATGTCGCCAGCTCTACCTGAGAAAGATGCTGTAGATTAACACAGACCTGTC 240	
Db 181 AGAAGTTCATGTCGCCAGCTCTACCTGAGAAAGATGCTGTAGATTAACACAGACCTGTC 240	
QY 241 TTCAGACCAACGCCCCGAGTTCAAGCTCTCTCCACACACTCAGGGAACAGATGTACTTGGCC 300	
Db 241 TTCAGACCAACGCCCCGAGTTCAAGCTCTCTCCACACACTCAGGGAACAGATGTACTTGGCC 300	
QY 301 CCGGCGACCGGAAACAGCTTTCAGGTTCAAGCTGACCTCGGGGACAGATGTCACTTGGCTC 360	
Db 301 CCGGCGACCGGAAACAGCTTTCAGGTTCAAGCTGACCTCGGGGACAGATGTCACTTGGCTC 360	
QY 361 CGAATTCACAGGTCACAGCTCTTGGGCTCTACACACCCCGACCCACGATGTCACTTGGCTC 420	
Db 361 CGAATTCACAGGTCACAGCTCTTGGGCTCTACACACCCCGACCCACGATGTCACTTGGCTC 420	

Qy	421	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	480
Dp	421	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	480
Qy	481	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	513
Dp	481	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	540
Qy	514	-----	-----	-----	-----	540
Dp	541	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	600
Qy	541	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	600
Dp	601	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	660
Qy	601	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	660
Dp	661	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	720
Qy	661	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	720
Dp	721	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	780
Qy	721	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	780
Dp	781	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	840
Qy	781	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	840
Dp	841	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	900
Qy	841	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	900
Dp	901	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	960
Qy	901	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	960
Dp	961	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1020
Qy	961	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1020
Dp	1021	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1080
Qy	1021	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1080
Dp	1081	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1140
Qy	1081	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1140
Dp	1141	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1200
Qy	1141	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1200
Dp	1201	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1260
Qy	1201	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1260
Dp	1261	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1320
Qy	1261	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1320
Dp	1321	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1380
Qy	1321	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1380
Dp	1381	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1440
Qy	1381	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1440
Dp	1441	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1500
Qy	1441	CCGGAACAAGCAGCCCGGAGCTCA	CCGCGCCCGCGGCTCA	CGGTGTA	CGTCACTCGGCG	1500


```
QY 1441 AGGATACCGATGTAAGCCCTATGAGAAGTTTTCGACGTAATGATGAGAGGCTTC 1500
DB 1501 AGGATACCGATGTAAGCCCTATGAGAAGTTTTCGACGTAATGATGAGAGGCTTC 1560
QY 1501 TCTTACAAACCCAGAGAGTGGACAGCACTTCTGCAACTTTGTAGAGGACAGTGGCCCTC 1560
DB 1561 TCTTACAAACCCAGAGAGTGGACAGCACTTCTGCAACTTTGTAGAGGACAGTGGCCCTC 1620
QY 1561 TGAGCTGAGTGG 1572
DB 1621 TGAGCTGAGTGG 1632

RESULT 8
US-10-029-517-3
; Sequence 3, Application US/10029517
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 3
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58) ... (1605)
US-10-029-517-3

Query Match 94.2%; Score 1481.2; DB 38; Length 1721;
Best Local Similarity 95.5%; Pred. No. 5.5e-280;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

QY 1 GAATTCCTGGGCTGCTGTAATCTGTCGCCCCCTCCGACCATTTTACACACCACTG 60
DB 1 GAATTCCTGGGCTGCTGTAATCTGTCGCCCCCTCCGACCATTTTACACACCACTG 60
QY 61 ACACCGGGGACCCAGTCTCTCTTCTCTCTGCTGCTCTCTCAAGTCTTACAGTTGTT 120
DB 61 ACACCGGGGACCCAGTCTCTCTTCTCTCTGCTGCTCTCTCAAGTCTTACAGTTGTT 120
QY 121 ACAGGTTCTGTATGATGCAAGCTCTACCCAGGTGAGAAAGAAAGAACTTGGCTACCCAG 180
DB 121 ACAGGTTCTGTATGATGCAAGCTCTACCCAGGTGAGAAAGAAAGAACTTGGCTACCCAG 180
QY 181 AGAAGTTCAAGTCCAGGCTCTACTGAGAAAGATGCTGTAGATGACACGACGCTATCTC 240
DB 181 AGAAGTTCAAGTCCAGGCTCTACTGAGAAAGATGCTGTAGATGACACGACGCTATCTC 240
QY 241 TCCAGCCACAGCCCGGTTCAAGGCTCTCCACACTCAGGAGACAGATGTCATCTGACC 300
DB 241 TCCAGCCACAGCCCGGTTCAAGGCTCTCCACACTCAGGAGACAGATGTCATCTGACC 300
QY 301 CCGGCGACCGGAAACAGGTTCAAGTTCAGCTGACACCTGGGAGACAGATGTCATCTGACC 360
DB 301 CCGGCGACCGGAAACAGGTTCAAGTTCAGCTGACACCTGGGAGACAGATGTCATCTGACC 360
QY 361 CCACTGACCAAGGCGGACCTTGGGCTCCACACCCCGCAGCCACAGATGTCATCTGACC 420
DB 361 CCACTGACCAAGGCGGACCTTGGGCTCCACACCCCGCAGCCACAGATGTCATCTGACC 420
QY 421 CCGGACAAACAGGCGGACCTTGGGCTCCACACCCCGCAGCCACAGATGTCATCTGACC 480
DB 421 CCGGACAAACAGGCGGACCTTGGGCTCCACACCCCGCAGCCACAGATGTCATCTGACC 480
QY 481 CCGGACAAACAGGCGGACCTTGGGCTCCACACCCCGCAGCCACAGATGTCATCTGACC 540
DB 481 CCGGACAAACAGGCGGACCTTGGGCTCCACACCCCGCAGCCACAGATGTCATCTGACC 540
```

```
QY 514 -----CCGCGGCCCAAGGTTTCACTTGGGCC 540
DB 541 CCGGACAAACAGGCGGACCTTGGGCTCCACACCCCGCAGCCACAGATGTCATCTGACC 600
QY 541 CCGGACAAACAGGCGGACCTTGGGCTCCACACCCCGCAGCCACAGATGTCATCTGACC 600
DB 601 CCGGACAAACAGGCGGACCTTGGGCTCCACACCCCGCAGCCACAGATGTCATCTGACC 660
QY 601 CCGGACAAACAGGCGGACCTTGGGCTCCACACCCCGCAGCCACAGATGTCATCTGACC 660
DB 661 CCGGACAAACAGGCGGACCTTGGGCTCCACACCCCGCAGCCACAGATGTCATCTGACC 720
QY 721 TGAGCTGATGATGAGGCTGATGTTGATCTGTGTCACAAAGGACCTTGGCGAGGCT 720
DB 721 TGAGCTGATGATGAGGCTGATGTTGATCTGTGTCACAAAGGACCTTGGCGAGGCT 780
QY 721 ACCAACAACCCAGGACGAGACATCCACAGATTTCCAGGACCACTCTGATTAAT 780
DB 781 ACCAACAACCCAGGACGAGACATCCACAGATTTCCAGGACCACTCTGATTAAT 840
QY 781 CTTACCAACCTTGGCAGCCATGACACCAAGATGATGCACTGATGACATGACAG 840
DB 841 CTTACCAACCTTGGCAGCCATGACACCAAGATGATGCACTGATGACATGACAG 900
QY 841 GTACTCTCTGACCTCTCCCAATCAAGCACTTCTCCAGTTGCTACTGAGGATCTCT 900
DB 901 GTACTCTCTGACCTCTCCCAATCAAGCACTTCTCCAGTTGCTACTGAGGATCTCT 960
QY 901 TTTCTTTTCTGCTCTTCTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCA 960
DB 961 TTTCTTTTCTGCTCTTCTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCA 1020
QY 961 AGCAGCACTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTCAGATTTAT 1020
DB 1021 AGCAGCACTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTCAGATTTAT 1080
QY 1021 AAAAAGAGGGGTTTCTGAGGCTCTCCCAATATTAATGACAGGACGATGAGTGGTGA 1080
DB 1081 AAAAAGAGGGGTTTCTGAGGCTCTCCCAATATTAATGACAGGACGATGAGTGGTGA 1140
QY 1081 CAATTGACTGTGCTCTTCCGAAAGATGATCCATGATGTCACAGAGTGGAGACAGTTTC 1140
DB 1141 CAATTGACTGTGCTCTTCCGAAAGATGATCCATGATGTCACAGAGTGGAGACAGTTTC 1200
QY 1141 AATCAGTATPAAAAGGAGAGAGCTCTGATATTAATGATGATGATGATGATGATGATG 1200
DB 1201 AATCAGTATPAAAAGGAGAGAGCTCTGATATTAATGATGATGATGATGATGATGATG 1260
QY 1201 AGTCAATGCTCATTTCTTTCTCTGCCCCAGTCTGGGATGCTGGGATGCTGGGATGCT 1260
DB 1261 AGTCAATGCTCATTTCTTTCTCTGCCCCAGTCTGGGATGCTGGGATGCTGGGATGCT 1320
QY 1261 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
DB 1321 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
QY 1321 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1381 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
QY 1381 ACTTACCATCTATGAGAGGATGACCTTACCAACCAACCAATGAGGCTGATGTCCTCT 1440
DB 1441 ACTTACCATCTATGAGAGGATGACCTTACCAACCAACCAATGAGGCTGATGTCCTCT 1500
QY 1441 AGCAGTACCGATGATGAGCCCTATGAGAAAGTTTTCAGAGTAAATGATGAGAGAGCT 1500
DB 1501 AGCAGTACCGATGATGAGCCCTATGAGAAAGTTTTCAGAGTAAATGATGAGAGAGCT 1560
QY 1501 TCTTACAAACCCAGAGAGTGGACAGCACTTCTGCAACTTTGTAGAGGACAGTGGCCCTC 1560
DB 1561 TCTTACAAACCCAGAGAGTGGACAGCACTTCTGCAACTTTGTAGAGGACAGTGGCCCTC 1620
QY 1561 TGAGCTGAGTGG 1572
```


[illegible]

```

RESULT 11
US-10-172-118-775
Sequence 775, Application US/10172118
GENERAL INFORMATION:
APPLICANT: Dai, Hongyue
APPLICANT: He, Yudong
APPLICANT: Linsley, Peter
APPLICANT: Mao, Mao
APPLICANT: Roberts, Chris
APPLICANT: Van 't Veer, Laura
APPLICANT: Van de Vijver, Marc
APPLICANT: Bernards, Rene
TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
FILE REFERENCE: 9301-175-999
CURRENT APPLICATION NUMBER: US/10/172, 118
CURRENT FILING DATE: 2002-06-14
PRIORITY APPLICATION NUMBER: 60/380, 770
PRIORITY FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 2699
SEQ ID NO 775
LENGTH: 1721
TYPE: DNA
ORGANISM: Homo sapiens
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: NM_002456
DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-775

```

Query Match	94.2%;	Score 1481.2;	DB 41,	Length 1721;
Best Local Similarity	95.5%;	Pred. No. 5,5e-280;		
Matches 1559; Conservative	0;	Mismatches 13;	Indels 60;	Gaps 1

Qy	1	GAATTCCTGGCGCTTGAATCGTTCGCGCCCTCCCCCAATTCAACACACATG	60
Db	1	GAATTCCTGGCGCTTGAATCGTTCGCGCCCTCCCCCAATTCAACACACATG	60
Qy	61	ACACCGGGCACCCAGTCTCTCTTCTCTCTGCTCTCTACAGTGCTTACAGTGT	120
Db	61	ACACCGGGCACCCAGTCTCTCTTCTCTCTGCTCTCTACAGTGCTTACAGTGT	120
Qy	121	ACAGGTTCTGGTCATGCAAGCTTACCCCAAGTGGAGAAAGAGACTTGGGCTACCCAG	180
Db	121	ACAGGTTCTGGTCATGCAAGCTTACCCCAAGTGGAGAAAGAGACTTGGGCTACCCAG	180
Qy	181	AGAAATTCAAGCGCCACAGCTTACATGAGAAAGAAATGCTGTGAGTAATGACACAGCTATCTC	240
Db	181	AGAAATTCAAGCGCCACAGCTTACATGAGAAAGAAATGCTGTGAGTAATGACACAGCTATCTC	240

Qy	241	TCGAGCACAAGCCCCGGTTTCAAGCTCTCCACA	CTCAGGAGACAGATGTCACCTGAGCC	300
Db	241	TCGAGCACAAGCCCCGGTTTCAAGCTCTCCACA	CTCAGGAGACAGATGTCACCTGAGCC	300
Qy	301	CCGAGCACAAGGACACAGCTTTCAGCTTGCCACA	CTGGGAGACAGATGTCACCTCGGTC	360
Db	301	CCGAGCACAAGGACACAGCTTTCAGCTTGCCACA	CTGGGAGACAGATGTCACCTCGGTC	360
Qy	361	CCAGTCAACAGGACCAAGCCTTGGGCTCCACA	CCCCGAGACCCACAGATGTCACCTCAAGCC	420
Db	361	CCAGTCAACAGGACCAAGCCTTGGGCTCCACA	CCCCGAGACCCACAGATGTCACCTCAAGCC	420
Qy	421	CCGAGCAAAAGAGCAGCCCGGGGCTCCACCG	CCCGCCCGCAGAGTGCACCTGAGCC	480
Db	421	CCGAGCAAAAGAGCAGCCCGGGGCTCCACCG	CCCGCCCGCAGAGTGCACCTGAGCC	480
Qy	481	CCGAGCACAGAGCCGCCCCGGGCTCCACCG	CCCGCCCGCAGAGTGCACCTGAGCC	540
Db	481	CCGAGCACAGAGCCGCCCCGGGCTCCACCG	CCCGCCCGCAGAGTGCACCTGAGCC	540
Qy	514	-----	-----CCCGCCCGCAGAGTGCACCTGAGCC	540
Db	541	CCGAGCACAGAGCCGCCCCGGGCTCCACCG	CCCGCCCGCAGAGTGCACCTGAGCC	600
Qy	541	CCGAGCACAGAGCCGCCCCGGGCTCCACCG	CCCGCCCGCAGAGTGCACCTGAGCC	600
Db	601	CCGAGCACAGAGCCGCCCCGGGCTCCACCG	CCCGCCCGCAGAGTGCACCTGAGCC	660
Qy	601	CCGAGCACAGAGCCGCCCCGGGCTCCACCG	CCCGCCCGCAGAGTGCACCTGAGCC	660
Db	661	CCGAGCACAGAGCCGCCCCGGGCTCCACCG	CCCGCCCGCAGAGTGCACCTGAGCC	720
Qy	661	CCGAGCACAGAGCCGCCCCGGGCTCCACCG	CCCGCCCGCAGAGTGCACCTGAGCC	720
Db	721	TCAGGCTCTGCAATCAGGCTCAGGCTCTTCA	CTTGGGAGACAAGGACCTCTCCAGAGGCT	780
Qy	721	TCAGGCTCTGCAATCAGGCTCAGGCTCTTCA	CTTGGGAGACAAGGACCTCTCCAGAGGCT	780
Db	781	ACGACACCCGACGACGAGAGAGCTCCACCG	ACATTCGACAGCACAACCTCGATGACT	840
Qy	781	ACGACACCCGACGACGAGAGAGCTCCACCG	ACATTCGACAGCACAACCTCGATGACT	840
Db	841	CTTACCAACCCCTTGCAGAGCATAGACCAAG	AGCTATGTCAGTAGCACAATACAGACG	900
Qy	841	CTTACCAACCCCTTGCAGAGCATAGACCAAG	AGCTATGTCAGTAGCACAATACAGACG	900
Db	901	GTACCTCTCTCAACCTCTCCATACAGACAT	TTCTCCCAAGTGTCTACACGGGCTCT	960
Qy	901	GTACCTCTCTCAACCTCTCCATACAGACAT	TTCTCCCAAGTGTCTACACGGGCTCT	960
Db	961	TTCTTTTCTCTGCTCTTTCACATTTCAAA	CTCCAGTTTAAATTCCTCTCTGAAAGATCCC	1020
Qy	961	TTCTTTTCTCTGCTCTTTCACATTTCAAA	CTCCAGTTTAAATTCCTCTCTGAAAGATCCC	1020
Db	1021	AGCACCGACCTTACCAAGAGCTGCAAGAG	CAATTTCTGAAATGTTTTTTCAGATTTAT	1080
Qy	1021	AGCACCGACCTTACCAAGAGCTGCAAGAG	CAATTTCTGAAATGTTTTTTCAGATTTAT	1080
Db	1081	AAACAAAGGGGTTTTCTGGGCTCTCCATAT	TAAAGTTTCAGGCGAGATCTGGAGTGGATA	1140
Qy	1081	AAACAAAGGGGTTTTCTGGGCTCTCCATAT	TAAAGTTTCAGGCGAGATCTGGAGTGGATA	1140
Db	1141	CAATTGACTCTGGCCTTCCAGAAAGGTAC	CAATATTCACAGAGTGGAGACACAGTTC	1200
Qy	1141	CAATTGACTCTGGCCTTCCAGAAAGGTAC	CAATATTCACAGAGTGGAGACACAGTTC	1200
Db	1201	AATCAGATATAAACGAGAGCAGCTCTCGAT	ATAACTGACATCTCAGAGTCAACGCG	1260
Qy	1201	AATCAGATATAAACGAGAGCAGCTCTCGAT	ATAACTGACATCTCAGAGTCAACGCG	1260
Db	1261	AGTATGTCGCAATTCCTTCTCTGCGCAG	GTGGAGTGGAGGCTCAGAGGCTGGAGATC	1320
Qy	1261	AGTATGTCGCAATTCCTTCTCTGCGCAG	GTGGAGTGGAGGCTCAGAGGCTGGAGATC	1320
Db	1321	GGGCTGCTGCTGCTGCTCTGTTGTTG	AGTGGAGGACCAATTCATCTCAATGCTCTTG	1380

Qy 1381 ACCGACACCTCCATGAGAGAGTACCCCAACCTACACACACAGAGGAGGCTATGAGCCCTT 1440
Db 1441 ACCGACACCTCCATGAGAGAGTACCCCAACCTACACACACAGAGGAGGCTATGAGCCCTT 1500
Qy 1441 AGCAGTACCGATCGTAGCCCTCTATGAGAAAGTTTCTGACAGTAAATGATGAGACAGCCCTC 1500
Db 1501 AGCAGTACCGATCGTAGCCCTCTATGAGAAAGTTTCTGACAGTAAATGATGAGACAGCCCTC 1560
Qy 1501 TCTTACACAAACCCAGCAGTGGAGCCACTTCTGACAACTTGTAGGAGGACAGTCCCTC 1560
Db 1561 TCTTACACAAACCCAGCAGTGGAGCCACTTCTGACAACTTGTAGGAGGACAGTCCCTC 1620
Qy 1561 TGAGCTGAGTGG 1572
Db 1621 TGAGCTGAGTGG 1632

RESULT 13
US-60-209-009-103
; Sequence 103, Application US/60209009
; GENERAL INFORMATION:
; APPLICANT: Paris, Mary
; APPLICANT: Pearson, Cecelia I.
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027 P
; CURRENT APPLICATION NUMBER: US/60/209,009
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL program
; SEQ ID NO: 103
; LENGTH: 2678
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 235943.39c
US-60-209-009-103

Query Match 94.2%; Score 1481.2; DB 64; Length 2678;
Best Local Similarity 95.5%; Pred. No. 5.8e-280;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

Qy 1 GAATTCCTGGCTGCTGATCTGTCTGCCCCCTCCCAACCATTTACACACACATG 60
Db 1 GAATTCCTGGCTGCTGATCTGTCTGCCCCCTCCCAACCATTTACACACACATG 60
Qy 61 ACACCGGGACCCAGATCTCTTTCTCTGCTGCTCTCTCAAGTGTTCAGAGTTGT 120
Db 61 ACACCGGGACCCAGATCTCTTTCTCTGCTGCTCTCTCAAGTGTTCAGAGTTGT 120
Qy 121 ACAGGTTCTGTCTGACAGCTCTACCCAGAGTGAAGAAAGAGACTTCGACTACCCAG 180
Db 121 ACAGGTTCTGTCTGACAGCTCTACCCAGAGTGAAGAAAGAGACTTCGACTACCCAG 180
Qy 181 AGAAGTTCAGGTCAGAGCTCTACTGAGAAAGATGTGTGATGATGACAGAGGTACTG 240
Db 181 AGAAGTTCAGGTCAGAGCTCTACTGAGAAAGATGTGTGATGATGACAGAGGTACTG 240
Qy 241 TCCAGCACAAGCCCGGTTCAAGGCTCTTCCACACTCAGGAGACAGATGTCACTCTGACC 300
Db 241 TCCAGCACAAGCCCGGTTCAAGGCTCTTCCACACTCAGGAGACAGATGTCACTCTGACC 300
Qy 301 CCGGACACGGAACCAAGTTCAAGTTCAAGTGCACCTGGGAGACAGATGTCACTCTGACC 360
Db 301 CCGGACACGGAACCAAGTTCAAGTTCAAGTGCACCTGGGAGACAGATGTCACTCTGACC 360
Qy 361 CCACTACACAGGACGAGCTGTGGCTCCACACACCCGACAGCCACAGATGTCACTCTGACC 420
Db 361 CCACTACACAGGACGAGCTGTGGCTCCACACACCCGACAGCCACAGATGTCACTCTGACC 420
Qy 421 CCGGAAACAAAGCCAGGCTCCAGGCTCCACCGGCTCCACCGGTCAACGTTCACTCTGACC 480
Db 421 CCGGAAACAAAGCCAGGCTCCAGGCTCCACCGGCTCCACCGGTCAACGTTCACTCTGACC 480

Db 421 CCGGAAACAAAGCCAGGCTCCAGGCTCCACCGGCTCCACCGGCTCCACCGGTCAACGTTCACTCTGACC 480
Qy 481 CCGGAAACAAAGCCAGGCTCCAGGCTCCACCGGCTCCACCGGCTCCACCGGTCAACGTTCACTCTGACC 513
Db 481 CCGGAAACAAAGCCAGGCTCCAGGCTCCACCGGCTCCACCGGCTCCACCGGTCAACGTTCACTCTGACC 540
Qy 514 -----CCGCGGCTCCAGGTTGTAACCTCGACC 540
Db 541 CCGGACACAGGCGGCTCCAGGCTCCACCGGCTCCACCGGCTCCACCGGTTGTAACCTCGACC 600
Qy 541 CCGGACACAGGCGGCTCCAGGCTCCACCGGCTCCACCGGCTCCACCGGTTGTAACCTCGACC 600
Db 541 CCGGACACAGGCGGCTCCAGGCTCCACCGGCTCCACCGGCTCCACCGGTTGTAACCTCGACC 660
Qy 601 CCGGAAACAAAGCCAGGCTCCAGGCTCCACCGGCTCCACCGGCTCCACCGGTTGTAACCTCGACC 660
Db 601 CCGGAAACAAAGCCAGGCTCCAGGCTCCACCGGCTCCACCGGCTCCACCGGTTGTAACCTCGACC 720
Qy 661 TCAAGCTCTGATCAGGCTTCAAGTTCTTACTCTGATGACAAAGGACCTCTGACAGGCT 720
Db 661 TCAAGCTCTGATCAGGCTTCAAGTTCTTACTCTGATGACAAAGGACCTCTGACAGGCT 780
Qy 721 ACCACAAACCCAGCAGCAGCAAGCACTCCACAGCATTCCACGACCACTCTGATCT 780
Db 721 ACCACAAACCCAGCAGCAGCAAGCACTCCACAGCATTCCACAGCACACTCTGATCT 840
Qy 781 CTTACACACCTTGGCAGCAGTACCAACAAAGTATGACAGTACCTACATAGACAG 840
Db 781 CTTACACACCTTGGCAGCAGTACCAACAAAGTATGACAGTACCTACATAGACAG 900
Qy 841 GTACCTCTCTCACTCTCTCCATACAGCACTCTCCGAGTTGTACTGAGGCTCT 900
Db 841 GTACCTCTCTCACTCTCTCCATACAGCACTCTCCGAGTTGTACTGAGGCTCT 960
Qy 901 TTCTTTTCTCTGCTTTTTCATATTCAAACTCCAGTTTATCTCTCTGAGATGCC 960
Db 901 TTCTTTTCTCTGCTTTTTCATATTCAAACTCCAGTTTATCTCTCTGAGATGCC 1020
Qy 961 AGCAGCAGTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTTTCAGATTAT 1020
Db 961 AGCAGCAGTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTTTCAGATTAT 1080
Qy 1021 AAACAAGGGGGTTTTCTGGGCTCTCCAAATTAAGTTACAGCAGAGATCTGTGTGTA 1080
Db 1021 AAACAAGGGGGTTTTCTGGGCTCTCCAAATTAAGTTACAGCAGAGATCTGTGTGTA 1140
Qy 1081 CAATTGACTTGGGCTTCCGAGAGGTACATCAATGTCCAGACGTGAGACACAGTTTC 1140
Db 1081 CAATTGACTTGGGCTTCCGAGAGGTACATCAATGTCCAGACGTGAGACACAGTTTC 1200
Qy 1141 AATCAGTATTAAGAGGAGAGCTCTCGATATTAACCTGACATCTCAGACGTGAGGTTG 1200
Db 1141 AATCAGTATTAAGAGGAGAGCTCTCGATATTAACCTGACATCTCAGACGTGAGGTTG 1260
Qy 1201 AGTCATGTCATTTCTCTTCTGTCAGTGTGGGCTGGGGTGCAGAGCTGGGGCATC 1260
Db 1201 AGTCATGTCATTTCTCTTCTGTCAGTGTGGGCTGGGGTGCAGAGCTGGGGCATC 1320
Qy 1261 GCGCTGCTGATGCTGATCTGTGTCTGATGCTGCTGAGCATTTGTCTATCTATGCTCTG 1320
Db 1261 GCGCTGCTGATGCTGATCTGTGTCTGATGCTGCTGAGCATTTGTCTATCTATGCTCTG 1380
Qy 1321 GCTGTCTGATGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1380
Db 1321 GCTGTCTGATGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1440
Qy 1381 GCTGTCTGATGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1440
Db 1381 GCTGTCTGATGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1500
Qy 1441 ACTTACATCTATGAGAGGATCCCACTTACACACACCAAGGAGGATGAGTGGCCCT 1500
Db 1441 ACTTACATCTATGAGAGGATCCCACTTACACACACCAAGGAGGATGAGTGGCCCT 1560
Qy 1501 AGCAGTACCGATCGTAGCCCTTATGAGAAAGTTTCTCAGAGTAAATGATGAGAGGCTC 1560
Db 1501 AGCAGTACCGATCGTAGCCCTTATGAGAAAGTTTCTCAGAGTAAATGATGAGAGGCTC 1560


```
/ Sequence 2117, Application US/09726172
/ GENERAL INFORMATION:
/ APPLICANT: Geating, David P.
/ APPLICANT: Holtzman, Douglas A.
/ TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
/ TITLE OF INVENTION: THEREFOR
/ FILE REFERENCE: 1600,2052-001
/ CURRENT APPLICATION NUMBER: US/09/726,172
/ PRIOR FILING DATE: 2000-11-29
/ PRIOR APPLICATION NUMBER: 60/167,846
/ NUMBER OF SEQ ID NOS: 2733
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2117
/ LENGTH: 2574
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ -09-726-172-2117

Query Match          92.5%; Score 1454.4; DB 29; Length 2574;
Best Local Similarity 95.3%; Pred. No. 18-274;
Matches 1534; Conservative 0; Mismatches 16; Indels 60; Gaps 1;

QY 1 GAATTCCTGCTGCTTGAATCTGTTCTGCCCCCTCCCACTTTCACACCACTCATG 60
DB 2572 GAATTCCTGCTGCTTGAATCTGTTCTGCCCCCTCCCACTTTCACACCACTCATG 2513

QY 61 ACACGGGGACCCAGCTCTCTTCTCTGCTGCTGCTCTCTCAAGTCTTACAGTTGTT 120
DB 2512 ACACGGGGACCCAGCTCTCTTCTCTGCTGCTGCTCTCTCAAGTCTTACAGTTGTT 2453

QY 121 ACAGTTTCTGTATGATGAAAGTCTACCCAGGTGAGAAAGAGACTTGGCTACCCAG 180
DB 2452 ACAGTTTCTGTATGATGAAAGTCTACCCAGGTGAGAAAGAGACTTGGCTACCCAG 2393

QY 181 AGAAGTTGAGTCCAGCTCTACTGAGAGAAATGCTGAGATGATGACCAAGATCTAC 240
DB 2392 AGAAGTTGAGTCCAGCTCTACTGAGAGAAATGCTGAGATGATGACCAAGATCTAC 2333

QY 241 TCCAGCCACAGCCCGGCTTCAAGGCTCTCTCCACCTCAAGGAGACAGATGCTAGTGC 300
DB 2332 TCCAGCCACAGCCCGGCTTCAAGGCTCTCTCTCCACCTCAAGGAGACAGATGCTAGTGC 2273

QY 301 CCGGACCAAGGAAACAGCTTCAAGTTGAGCTGCAAGCTGAGGAGACAGATGCTAGTGC 360
DB 2272 CCGGACCAAGGAAACAGCTTCAAGTTGAGCTGCAAGCTGAGGAGACAGATGCTAGTGC 2213

QY 361 CCAATGACCAAGGCAAGCTTGGAGCTTCCACCAAGCTGCAAGCTGCAAGCTTGGAGCT 420
DB 2212 CCAATGACCAAGGCAAGCTTGGAGCTTCCACCAAGCTGCAAGCTGCAAGCTTGGAGCT 2153

QY 421 CCGGACCAAGGCAAGCTTGGAGCTTCCACCAAGCTGCAAGCTGCAAGCTTGGAGCT 480
DB 2152 CCGGACCAAGGCAAGCTTGGAGCTTCCACCAAGCTGCAAGCTGCAAGCTTGGAGCT 2093

QY 481 CCGGACCAAGGCAAGCTTGGAGCTTCCACCAAGCTGCAAGCTTGGAGCT 513
DB 2092 CCGGACCAAGGCAAGCTTGGAGCTTCCACCAAGCTGCAAGCTTGGAGCT 2033

QY 514 -----CCGGCGGCGCAAGGTTGCAAGCTTGGAGCT 540
DB 2032 CCGGACCAAGGCAAGCTTGGAGCTTCCACCAAGCTGCAAGCTTGGAGCT 1973

QY 541 CCGGACCAAGGCAAGCTTGGAGCTTCCACCAAGCTGCAAGCTTGGAGCT 600
DB 1972 CCGGACCAAGGCAAGCTTGGAGCTTCCACCAAGCTGCAAGCTTGGAGCT 1913

QY 601 CCGGACCAAGGCAAGCTTGGAGCTTCCACCAAGCTGCAAGCTTGGAGCT 660
DB 1912 CCGGACCAAGGCAAGCTTGGAGCTTCCACCAAGCTGCAAGCTTGGAGCT 1853

QY 661 TGAAGCTCTGATGAGGCTGAGCTTCTACTTGTGTGCAAAAGGCACTTGGCAAGGCT 720
DB 1853 TGAAGCTCTGATGAGGCTGAGCTTCTACTTGTGTGCAAAAGGCACTTGGCAAGGCT 720
```

```
DB 1852 TCAGGCTCTGATGAGGCTGAGCTTCTACTTGTGTGCAAAAGGCACTTGGCAAGGCT 1793
QY 721 ACACCAACCCCAAGCCAGCAAGGCTCTCCACCAAGCTTCCAGCCACCACTGATATCT 780
DB 1792 ACACCAACCCCAAGCCAGCAAGGCTCTCCACCAAGCTTCCAGCCACCACTGATATCT 1733
QY 781 CCTACCAACCCCTGAGCCATGAGCAAGGCTGATGAGCACTTACCATAGCAGC 840
DB 1732 CCTACCAACCCCTGAGCCATGAGCAAGGCTGATGAGCACTTACCATAGCAGC 1673
QY 841 GTACTCTCTTACCTCTCTCCATATCAAGCACTTCTCCCAAGTGTCTACTGGAGTCTCT 900
DB 1672 GTACTCTCTTACCTCTCTCCATATCAAGCACTTCTCCCAAGTGTCTACTGGAGTCTCT 1613
QY 901 TTTCTTTTCTGCTCTTTTCAATTTCAATTTCAATTTCAATTTCTCTCTGGAAGATCCC 960
DB 1612 TTTCTTTTCTGCTCTTTTCAATTTCAATTTCAATTTCAATTTCTCTCTGGAAGATCCC 1553
QY 961 AGCACCGACTACTACCAAGAGCTGCAAGAGACATTTCTGAAATGTTTTGCAATTTAT 1020
DB 1552 AGCACCGACTACTACCAAGAGCTGCAAGAGACATTTCTGAAATGTTTTGCAATTTAT 1493
QY 1021 AAACAAAGGGGTTTTCTGGGCTCTCCAAATTAAGTTCAAGGCAAGATCTGTGGTGA 1080
DB 1492 AAACAAAGGGGTTTTCTGGGCTCTCCAAATTAAGTTCAAGGCAAGATCTGTGGTGA 1433
QY 1081 CAATTAAGCTTGGCTCTTCCGAAAGATGATCAATGTTCCAGAGTGGAGACACTTTC 1140
DB 1432 CAATTAAGCTTGGCTCTTCCGAAAGATGATCAATGTTCCAGAGTGGAGACACTTTC 1373
QY 1141 AATCAGATTAACCAAGGAGCAGCTCTGATATTAACCTGACATCTCAGAGTCAAGCTG 1200
DB 1372 AATCAGATTAACCAAGGAGCAGCTCTGATATTAACCTGACATCTCAGAGTCAAGCTG 1313
QY 1201 AGTCATGTCATTTCTCTTCTCTGCCCCAGTCTGGGGTCCAGGCTGGGGATC 1260
DB 1312 AGTCATGTCATTTCTCTTCTCTGCCCCAGTCTGGGGTCCAGGCTGGGGATC 1253
QY 1261 GCGCTGTGCTGCTGCTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
DB 1252 GCGCTGTGCTGCTGCTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1193
QY 1321 GCTGTCTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1192 GCTGTCTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1133
QY 1381 ACCTACCATCTATAGAGGAGTACCCCACTTACCAACCCATGAGGCTGATGCTCCCT 1440
DB 1132 ACCTACCATCTATAGAGGAGTACCCCACTTACCAACCCATGAGGCTGATGCTCCCT 1073
QY 1441 AGCAGTACCGATGCTGAGCCCTATGAGAAAGTCTTGGAGGTAAGTGTAGAGAGCTC 1500
DB 1072 AGCAGTACCGATGCTGAGCCCTATGAGAAAGTCTTGGAGGTAAGTGTAGAGAGCTC 1013
QY 1501 TCTTAACAACCAAGCAGAGTGGAGCCACTTGGCCACTTGTGTAGGGGCA 1550
DB 1012 TCTTAACAACCAAGCAGAGTGGAGCCACTTGGCCACTTGTGTAGGGGCA 963
```

Search completed: May 8, 2003, 18:01:38
Job time : 3769 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:14:01 ; Search time 74 Seconds
(without alignments)

16.206 Million cell updates/sec

Title: US-09-658-621B-26

Perfect score: 44

Sequence: 1 ISEMFLQIY 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 130868

Minimum DB seq length: 0

Maximum DB seq length: 9

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq_101002:*

1: /SID82/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SID82/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SID82/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SID82/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SID82/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SID82/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SID82/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SID82/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SID82/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SID82/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SID82/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SID82/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SID82/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SID82/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SID82/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SID82/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SID82/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SID82/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SID82/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SID82/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SID82/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SID82/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SID82/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	44	100.0	9	22	AAU00563	Human MUC1 polypep
2	22	50.0	9	19	AAAS9466	HIV-1 Rev protein
3	22	50.0	9	19	AAAS9469	HIV-1 Rev protein
4	22	50.0	9	21	AAV58724	Paraneoplastic cer
5	22	50.0	9	22	AAAG9553	p53 epitope A2 sup
6	22	50.0	9	22	AAAG9585	p53 epitope A2 sup
7	22	50.0	9	22	AAAG9706	p53 epitope HLA-A2
8	22	50.0	9	23	AAU95955	Immunogenic peptid
9	21	47.7	8	16	AAAR73578	Human TSH receptor
10	21	47.7	8	21	AAAB05855	Peptide identified

11	21	47.7	9	14	AAAR4265	Residues 129-137 o
12	21	47.7	9	14	AAAR4282	Residues 129-137 o
13	21	47.7	9	15	AAAY38357	PAP-derived HLA-b1
14	21	47.7	9	15	AAAR59189	Peptide Fragment (
15	21	47.7	9	17	AAAR97530	Antigenic peptide,
16	21	47.7	9	20	AAAY53361	p53 epitope (aa 12
17	21	47.7	9	20	AAAY53361	Immunogenic peptid
18	21	47.7	9	20	AAAY40184	Amino acid sequenc
19	21	47.7	9	20	AAAY26700	HLA-A2 binding p53
20	21	47.7	9	21	AAAB23704	Cytotoxic T lympho
21	21	47.7	9	21	AAAB23704	Cytotoxic T lympho
22	21	47.7	9	22	AAAG67192	Cancer testis tumo
23	21	47.7	9	22	AAAG89424	p53 DR supermotif
24	21	47.7	9	22	AAAG89584	p53 epitope A2 sup
25	21	47.7	9	22	AAAG89584	p53 epitope A2 sup
26	21	47.7	9	22	AAAG89705	p53 epitope HLA-A2
27	21	47.7	9	22	AAAB69924	p53 epitope HLA-A2
28	21	47.7	9	22	AAAB76012	Human NY-ESO-1 HLA
29	20	45.5	5	12	AAAR13643	Tumour associated
30	20	45.5	5	12	AAAR13646	HIV protease inhib
31	20	45.5	7	22	AAAR46244	HIV protease inhib
32	20	45.5	7	22	AAAR46249	H11 binding site c
33	20	45.5	7	22	AAAR46378	H11 binding site c
34	20	45.5	7	22	AAAR46378	H11 binding site c
35	20	45.5	8	15	AAAR70990	H11 binding site c
36	20	45.5	9	15	AAAY38031	Bovine ET receptor
37	20	45.5	9	15	AAAY38032	Hepatitis B virus-
38	20	45.5	9	20	AAAY45601	Hepatitis B virus-
39	20	45.5	9	20	AAAY03004	Immunogenic peptid
40	20	45.5	9	22	AAAR05206	Fragment of human
41	20	45.5	9	23	AAU94465	Human HLA-A1 bindi
42	20	45.5	9	23	AAU94482	Human novel protei
43	19	43.2	5	12	AAAR13644	Human novel protei
44	19	43.2	5	12	AAAR13645	HIV protease inhib
45	19	43.2	6	11	AAAR08214	HIV protease and r

ALIGNMENTS

RESULT 1	AAU00563	standard; Peptide; 9 AA.
ID	AAU00563	
AC	AAU00563;	
XX		
DT	12-SEP-2001	(first entry)
XX		
DE	Human MUC1 polypeptide derivative #24.	
XX		
KW	Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-I;	
KW	glycoprotein; cytotoxic T lymphocytes; T cell response; cancer; vaccine;	
KW	cancer gene therapy; diagnosis; treatment; inflammatory disorder;	
KW	organ transplant rejection; graft versus host disease.	
OS	Homo sapiens.	
XX		
PN	WO200118035-A2.	
XX		
PD	15-MAR-2001.	
XX		
PF	07-SEP-2000; 2000WO-EP08761.	
XX		
PR	08-SEP-1999; 99GB-0021242.	
PR	10-SEP-1999; 99EP-0402237.	
PR	03-MAR-2000; 2000US-0187215.	
XX		
PA	(TRGS) TRANSGENE SA.	
XX	(IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.	
PI	Taylor-Papadimitriou J, Heukamp LC, Offringa R, Melief CJM;	
XX	Acree B, Thomas M;	
XX		

DR WPI: 2001-235187/24.
DR N-PSDB; AAS00609.

PT New antigenic polypeptides of MUC-1 protein which activate cytotoxic T
PT lymphocyte proteins and their analogues, useful for identifying a major
PT histocompatibility complex class I restricted T cell response and for
PT diagnosing cancer

PS Claim 2; Page 74; 81pp; English.

CC The sequence represents a human MUC1 polypeptide derivative. Derivative
CC antigenic peptides of MUC1 protein bind at least one major
CC histocompatibility complex class I (MHC-I) glycoprotein, which activates
CC cytotoxic T lymphocytes to induce a protective response against tumours.
CC Diagnosis of cancer involves determining the presence or absence in a
CC host cell of MHC class I restricted T cell response to a MUC1 derivative,
CC where the presence of the MHC class I restricted T cell response
CC indicates that the host has cancer. Measurement of the level of MHC class
CC I restricted T cell response is also useful to monitor the severity of
CC cancer, a larger response indicating a more severe cancer. MUC1
CC derivatives are useful in cancer therapy and to follow MUC1 specific
CC immune responses in patients during the course of disease and/or
CC treatment. MUC1 DNA is useful in cancer gene therapy, vaccination and
CC diagnosis. Compositions of the sequences are used in vaccines and
CC treatments against cancer or diseases caused by an immune response, such
CC as an inflammatory disorder, organ transplant rejection or graft versus
CC host disease.

SO Sequence 9 AA;

Query Match 100.0%; Score 44; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMFLQIY 9
DB 1 ISEMFLQIY 9

RESULT 2

AAM59466
ID AAM59466 standard; peptide; 9 AA.

AAW59466;

28-AUG-1998 (first entry)

HIV-1 Rev protein HLA-A1 peptide motif #1 from patient L658.

Human immunodeficiency virus; HIV; AIDS; cytotoxic T cell; lymphocyte;
acquired immunodeficiency syndrome; Rev protein; epitope; immunisation;
detection; prognosis; diagnosis; treatment; prevention.

Human immunodeficiency virus type 1.

WO9817309-A1.

30-APR-1998.

17-OCT-1997; 97WO-IB01402.

18-OCT-1996; 96US-0733789.

(UVR0-) UNIV ROTTERDAM ERASMUS.

Osterhaus ADME, Van Baalen CA;

WPI; 1998-261189/23.

Use of Rev or Tat proteins of immunodeficiency virus - for
developing products for diagnosis, prognosis, prevention or
treatment of infection, especially HIV infection

PS Disclosure; Page 19; 38pp; English.

CC AAM59466-W59480 are fragments of the human immunodeficiency virus Rev
CC protein which are used in a novel method to detect at least 1 cytotoxic
CC protein epitope of the virus in order to prevent the disease. The
CC specification also describes a method of immunising a host against
CC disease caused by infection by an immunodeficiency virus, which
CC comprises stimulating in a host a specific cytotoxic T-cell response
CC which is specific for the Rev and/or Tat proteins of the immunodeficiency
CC virus. A method determining favourable prognosis in an HIV positive
CC subject which comprises detecting the presence of a cytotoxic T-cell
CC response to Rev and/or Tat HIV protein as an indication of the favourable
CC the diagnosis, monitoring, prevention and treatment of immunodeficiency
CC disease mediated by an immunodeficiency virus, especially HIV.

SO Sequence 9 AA;

Query Match 50.0%; Score 22; DB 19; Length 9;
Best Local Similarity 55.6%; Pred. No. 7.8e+05;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ISEMFLQIY 9
DB 1 ISEIRLSTY 9

RESULT 3

AAM59469
ID AAM59469 standard; peptide; 9 AA.

AAW59469;

28-AUG-1998 (first entry)

HIV-1 Rev protein HLA-A1 peptide motif #1 from patient P424.

Human immunodeficiency virus; HIV; AIDS; cytotoxic T cell; lymphocyte;
acquired immunodeficiency syndrome; Rev protein; epitope; immunisation;
detection; prognosis; diagnosis; treatment; prevention.

Human immunodeficiency virus type 1.

WO9817309-A1.

30-APR-1998.

17-OCT-1997; 97WO-IB01402.

18-OCT-1996; 96US-0733789.

(UVR0-) UNIV ROTTERDAM ERASMUS.

Osterhaus ADME, Van Baalen CA;

WPI; 1998-261189/23.

Use of Rev or Tat proteins of immunodeficiency virus - for
developing products for diagnosis, prognosis, prevention or
treatment of infection, especially HIV infection

Disclosure; Page 19; 38pp; English.

CC AAM59466-W59480 are fragments of the human immunodeficiency virus Rev
CC protein which are used in a novel method to detect at least 1 cytotoxic
CC protein epitope of the virus in order to prevent the disease. The
CC specification also describes a method of immunising a host against
CC disease caused by infection by an immunodeficiency virus, which
CC comprises stimulating in a host a specific cytotoxic T-cell response
CC which is specific for the Rev and/or Tat proteins of the immunodeficiency
CC virus. A method determining favourable prognosis in an HIV positive
CC subject which comprises detecting the presence of a cytotoxic T-cell
CC response to Rev and/or Tat HIV protein as an indication of the favourable

CC present invention.

XX Sequence 9 AA;

Query Match 50.0%; Score 22; DB 22; Length 9;
Best Local Similarity 37.5%; Pred. No. 7.8e+05;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEWFLQI 8
:::|:::
DB 2 LNKMFQCV 9

RESULT 6

AAAG89585
ID AAG89585 standard; Peptide; 9 AA.

XX AAG89585;

DT 11-SEP-2001 (first entry)

DE p53 epitope A2 supermotif crossbinding peptide #20.

XX Human; p53; tumour suppressor; human leukocyte antigen; HLA; CTL;
XX cytotoxic T lymphocyte; T cell; immune response; tumour; cancer;
XX vaccine; epitope; cytostatic.

OS Homo sapiens.

OS Synthetic.

XX WO200141788-A1.

XX 14-JUN-2001.

XX 11-DEC-2000; 2000WO-US33629.

XX 10-DEC-1999; 99US-0458297.

XX (EPIM-) EPIMUNE INC.

XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX Keogh E;

XX WPI; 2001-381493/40.

PT Epitope-based vaccines comprising p53 epitope having a specified
PT sequences, useful for treating and preventing cancer, the epitopic
PT peptides is useful as diagnostic agents and for evaluating immune
PT response

PS Example 4; Page 120; 138pp; English.

CC The present invention describes isolated prepared p53 epitopes (I). Also
CC described are: (1) a cytotoxic T lymphocyte (CTL) that is cultured
CC in vivo and binds to a complex of (I); (2) a peptide (II) comprising (I)
CC and a second epitope and has less than 50 contiguous amino acids; (3) a
CC vaccine composition comprising (II), a unit dose of a peptide with less
CC than 50 contiguous amino acids with 100% identity to the native peptide
CC sequence of p53; and a pharmaceutical excipient; (4) an isolated nucleic
CC acid encoding (I); and (5) an isolated nucleic acid encoding (II). (I)
CC has cytostatic activity and can be used in vaccines. The vaccine
CC composition is useful for treating or preventing cancer. (I) and (II)
CC are useful as diagnostic agents and for evaluating immune responses.
CC Unlike conventional epitopes, immunosuppressive epitopes that may be
CC present in whole antigens can be avoided with the use of the vaccine
CC composition of (I). The ability to combine selected epitopes and
CC further, to modify the composition of the epitopes enhances the
CC immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigens, which might have their own
CC intrinsic biological activity, are eliminated. AAG89363 to AAG89747
CC represent amino acid sequences used in the exemplification of the
CC present invention.

SQ Sequence 9 AA;

Query Match 50.0%; Score 22; DB 22; Length 9;
Best Local Similarity 37.5%; Pred. No. 7.8e+05;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEWFLQI 8
:::|:::
DB 2 LNKMFQCV 9

RESULT 7

AAAG89706
ID AAG89706 standard; Peptide; 9 AA.

XX AAG89706;

DT 11-SEP-2001 (first entry)

DE p53 epitope HLA-A2 supermotif peptide #5.

XX Human; p53; tumour suppressor; human leukocyte antigen; HLA; CTL;
XX cytotoxic T lymphocyte; T cell; immune response; tumour; cancer;
XX vaccine; epitope; cytostatic.

OS Homo sapiens.

OS Synthetic.

XX WO200141788-A1.

XX 14-JUN-2001.

XX 11-DEC-2000; 2000WO-US33629.

XX 10-DEC-1999; 99US-0458297.

XX (EPIM-) EPIMUNE INC.

XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX Keogh E;

XX WPI; 2001-381493/40.

PT Epitope-based vaccines comprising p53 epitope having a specified
PT sequences, useful for treating and preventing cancer, the epitopic
PT peptides is useful as diagnostic agents and for evaluating immune
PT response

PS Claim 1; Page 128; 138pp; English.

CC The present invention describes isolated prepared p53 epitopes (I). Also
CC described are: (1) a cytotoxic T lymphocyte (CTL) that is cultured
CC in vivo and binds to a complex of (I); (2) a peptide (II) comprising (I)
CC and a second epitope and has less than 50 contiguous amino acids; (3) a
CC vaccine composition comprising (II), a unit dose of a peptide with less
CC than 50 contiguous amino acids with 100% identity to the native peptide
CC sequence of p53; and a pharmaceutical excipient; (4) an isolated nucleic
CC acid encoding (I); and (5) an isolated nucleic acid encoding (II). (I)
CC has cytostatic activity and can be used in vaccines. The vaccine
CC composition is useful for treating or preventing cancer. (I) and (II)
CC are useful as diagnostic agents and for evaluating immune responses.
CC Unlike conventional epitopes, immunosuppressive epitopes that may be
CC present in whole antigens can be avoided with the use of the vaccine
CC composition of (I). The ability to combine selected epitopes and
CC further, to modify the composition of the epitopes enhances the
CC immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigens, which might have their own
CC intrinsic biological activity, are eliminated. AAG89363 to AAG89747
CC represent amino acid sequences used in the exemplification of the
CC present invention.

XX Sequence 9 AA;

Query Match 50.0%; Score 22; DB 22; Length 9;
 Best Local Similarity 37.5%; Pred. No. 7.8e+05;
 Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 8
 DB 2 LNMFCOV 9

RESULT 8

AAU95955

ID AAU95955 standard; Peptide; 9 AA.

AC AAU95955;

DT 02-JUL-2002 (first entry)

DE Immunogenic peptide with (HLA)-A2.1 binding site #168.

XX HLA-A2.1 binding peptide; cytostatic; virucide; anti-HIV; hepatotropic;
 XX human immunodeficiency virus; anti-inflammatory; antibacterial; vaccine;
 XX protozoacide; immunosuppressant; immunogenic peptide; T cell activation;
 XX human leucocyte antigen binding site; cytotoxic T cell response;
 XX viral infection; hepatitis; Epstein-Barr virus; papilloma virus;
 XX human immunodeficiency virus; HIV; Kaposi sarcoma; Lassa fever virus;
 XX cytomegalovirus; tumour; prostate cancer; renal carcinoma; lymphoma;
 XX prostate-specific antigen; p53; carcino-embryonal antigen;
 XX melanoma antigen; Mycobacterium tuberculosis; protozoa;
 XX trypanosome surface antigen; condyloma acuminatum.

XX Unidentified.

OS WO200220616-A1.

PN 14-MAR-2002.

PD 01-SEP-2000; 2000MO-US24102.

PF 01-SEP-2000; 2000MO-US24102.

PR 01-SEP-2000; 2000MO-US24102.

XX (EPIM-) EPIMUNE INC.

PA Grey HM, Sette A, Sidney J, Southwood S;

PI WPI; 2002-351766/38.

XX Immunogenic peptide with human leucocyte antigen-A2.1 binding site,

PT useful for treating e.g. viral infection or tumours

XX Claim 1; Page 29; 35pp; English.

XX The invention describes a composition comprising an immunogenic peptide
 CC having a human leucocyte antigen (HLA)-A2.1 binding site. The peptides
 CC bind specifically to HLA-A2.1, to cause T cell activation and thus a
 CC cytotoxic T cell response. The peptides and the nucleic acids that
 CC encode them, are used, in vivo or ex vivo, for treatment of viral
 CC infections (hepatitis B or C; Epstein-Barr; human immune deficiency;
 CC Kaposi sarcoma; human papilloma; Lassa fever or cytomegalovirus);
 CC tumours including prostate cancer, renal carcinoma and lymphoma (where
 CC directed to prostate-specific antigen, p53, carcino-embryonal antigen,
 CC Her2/neu or melanoma antigens); infection by Mycobacterium tuberculosis
 CC or protozoa (directed to trypanosome surface antigen); and condyloma
 CC acuminatum. The peptides are suitable for use in peptide-based
 CC vaccines. This sequence represents an immunogenic peptide with the
 CC human leucocyte antigen (HLA)-A2.1 binding site, described in the
 CC invention.

XX Sequence 9 AA;

Query Match 50.0%; Score 22; DB 23; Length 9;
 Best Local Similarity 37.5%; Pred. No. 7.8e+05;
 Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 8
 DB 2 LNMFCOV 9

RESULT 9

AAR73578

ID AAR73578 standard; Peptide; 8 AA.

AC AAR73578;

DT 19-DEC-1995 (first entry)

DE Human TSH receptor (residues 755-762).

XX thyroid stimulating hormone receptor; TSH; human; Homo sapiens;
 XX antibody; affinity; detection.

XX Synthetic.

PN JP07089991-A.

XX 04-APR-1995.

PD 28-SEP-1993; 93JP-0240853.

PF 28-SEP-1993; 93JP-0240853.

PR (MITP) MITSUBISHI PETROCHEMICAL CO LTD.

XX WPI; 1995-167251/22.

XX Novel polypeptide(s) having affinity for the human TSH receptor
 PT antibody - used in detection of the TSH antibody.

XX Example 1; Page 51; 54pp; Japanese.

XX Peptides with affinity to human TSH (thyroid stimulating hormone)
 CC receptor antibody are used for detection of the antibody. (See also
 CC AAR73201-592).

XX Sequence 8 AA;

Query Match 47.7%; Score 21; DB 16; Length 8;
 Best Local Similarity 57.1%; Pred. No. 7.8e+05;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 7
 DB 1 ISEMFLOI 7

RESULT 10

ID AAB08585 standard; Peptide; 8 AA.

AC AAB08585;

DT 20-DEC-2000 (first entry)

DE Peptide identified from an origin of DNA polymerase epsilon.

XX Precursor peptide; polypeptide hormone; peptide identification.

XX Unidentified.

XX Key Location/Qualifiers

XX Modified-site 1 /note= "hydrogen attached"

XX Modified-site 8 /note= "amidated residue"

XX WO200050636-A1.

```

XX 31-AUG-2000.
PD
XX
XX 24-FEB-2000; 2000MO-FR00460.
PF
XX 25-FEB-1999; 99US-0257525.
PR
XX (SCRC ) SCRAS SOC CONSEILS RECH & APPL SCT.
PA (CNRS ) CNRS CENT NAT RECH SCT.
XX
XX Camara Ferrer YJA, Thuriereau C, Martinez J, Borge G, Goze C;
XX WPI; 2000-572101/53.
DR
XX
XX Identifying peptide with selected function, useful particularly for
PT C-amidated hormones, by screening database for combination of nucleic
PT acid and amino acid sequences
XX
XX Disclosure; Page 21; 40pp; French.
PS
XX The specification describes a method for identifying a peptide having
CC a particular function. The method comprises preparing a database of
CC polynucleotides and polypeptides of unknown functions, screening the
CC database for a combination of nucleotides or amino acids indicative of
CC the peptide with a particular function, and identifying polynucleotides
CC and proteins which contain the peptide. The method is used to identify
CC precursor peptides with an amidated C-terminus, especially polypeptide
CC hormones, for studying physiologically active substances. The present
CC sequence represents a peptide which was identified using the method of
CC the invention.
XX
XX Sequence 8 AA;
SQ
XX
XX Query Match 47.7%; Score 21; DB 21; Length 8;
XX Best Local Similarity 80.0%; Pred. No. 7.8e+05;
XX Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 ISEMF 5
Db 1 ISELF 5
XX
XX RESULT 11
XX AAR44265
XX ID AAR44265 standard; peptide; 9 AA.
XX
XX AAR44265;
XX 09-JUN-1994 (first entry)
XX
XX Residues 129-137 of human p53.
XX
XX Human; p53; protein; Class I; allele; diagnosis; assay; detection;
XX p53 protein-specific; T-cell; antibody; cancer; MHC; HLA-A2.1.
XX
XX Homo sapiens.
XX
XX WO9324525-A.
XX
XX 09-DEC-1993.
XX
XX 18-MAY-1993; 93WO-NL00102.
XX
XX 26-MAY-1992; 92EP-0201510.
XX
XX (UYLE-) RIJKSUNIV LEIDEN.
XX (SEED-) SEED CAPITAL INVESTMENTS (SCI) BV.
XX
XX Kast WM, Melief CJM;
XX WPI; 1993-405730/50.
XX
XX Peptide derived from p53 protein - used to treat and diagnose
XX peptide derived from p53 protein - used to treat and diagnose
PT

```

```

PT diseases involving over-expression of p53 e.g. human cancers
XX
XX Claim 3; Page 37; 46pp; English.
PS
XX
XX The sequences given in AAR44265-80 are peptides derived from human p53
CC protein which have the ability to bind to a human Class I molecule.
CC These peptides may be used in a diagnostic test or assay to detect
CC human p53 protein-specific T-cells or antibodies. They may also be
CC used in the treatment of diseases such as human cancers showing p53
CC protein overexpression. These peptides can bind human MHC Class I
XX allele HLA-A2.1.
XX
XX Sequence 9 AA;
SQ
XX
XX Query Match 47.7%; Score 21; DB 14; Length 9;
XX Best Local Similarity 37.5%; Pred. No. 7.8e+05;
XX Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 ISEMFLOI 8
Db 2 INKMFCDL 9
XX
XX RESULT 12
XX AAR44282
XX ID AAR44282 standard; peptide; 9 AA.
XX
XX AAR44282;
XX
XX 09-JUN-1994 (first entry)
XX
XX Residues 129-137 of human mp53.
XX
XX Human; p53; protein; Class I; allele; diagnosis; assay; detection;
XX p53 protein-specific; T-cell; antibody; cancer; MHC; HLA-A2.1.
XX
XX Homo sapiens.
XX
XX WO9324525-A.
XX
XX 09-DEC-1993.
XX
XX 18-MAY-1993; 93WO-NL00102.
XX
XX 26-MAY-1992; 92EP-0201510.
XX
XX (UYLE-) RIJKSUNIV LEIDEN.
XX (SEED-) SEED CAPITAL INVESTMENTS (SCI) BV.
XX
XX Kast WM, Melief CJM;
XX WPI; 1993-405730/50.
XX
XX Peptide derived from p53 protein - used to treat and diagnose
XX diseases involving over-expression of P53 e.g. human cancers
XX
XX Claim 3; Page 38; 46pp; English.
XX
XX The sequences given in AAR44281-94 are peptides derived from human mp53
CC protein which have the ability to bind to a human Class I molecule.
CC These peptides may be used in a diagnostic test or assay to detect
CC human p53 protein-specific T-cells or antibodies. They may also be
CC used in the treatment of diseases such as human cancers showing p53
CC protein overexpression. These peptides can bind human MHC Class I
XX allele HLA-A2.1.
XX
XX Sequence 9 AA;
SQ
XX
XX Query Match 47.7%; Score 21; DB 14; Length 9;
XX Best Local Similarity 37.5%; Pred. No. 7.8e+05;
XX Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 ISEMFLOI 8

```



```

PR 14-DEC-1994; 9AUS-0355558.
XX
XX (SCRI ) SCRIIPS RES INST.
XX
XX Sherman LA;
XX
XX WPI; 1996-300385/30.
XX
XX In vivo activation of tumour-specific cytotoxic T lymphocytes - by
PT contacting with polypeptide(s) derived from human p53 or Her-2/Neu
PT proteins
XX
XX Example 4; Page 94; 158pp; English.
XX
CC AA97517-R97544 are antigenic p53-derived peptides synthesised to be
CC used in an assay for identifying peptides capable of activating
CC cytotoxic T lymphocytes (CTLs) which specifically target malignant
CC cells. CTL-activating peptides can be used in a vaccine for
CC protecting against tumour cell formation. CTLs activated by the
CC peptides will lyse tumour cells displaying specific peptides.
CC Antibodies against CTL-activating peptides are useful for the
CC identification of other similar compounds which may be useful for
CC treating cancer or virally-infected cells, or for diagnosis. The
CC peptide and vaccines produced provide immunity to a high percentage
CC of different ethnic groups, i.e. those with different HLA alleles.
XX
SQ Sequence 9 AA;

Query Match 47.7%; Score 21; DB 17; Length 9;
Best Local Similarity 37.5%; Pred. No. 7.8e+05;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 ISEMFLQI 8
   ::::|:
   2 INKMECOL 9
   Db

```

Search completed: May 1, 2003, 23:22:57
 Job time : 75 secs

Thu May, 8 18:53:07 2003

us-09-658-621b-1.rnt

Page 1

GenCore version 5.1.4.D5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic ~ nucleic search, using sw model

Run on: May 8, 2003, 14:56:22 ; Search time 73 Seconds
(without alignments)
6604.053 Million cell updates/sec

Title: US-09-658-621B-1

Perfect score: 1572
Sequence: 1 gattccctgagctgtgaa.....tcgcccctgagctgagrrg 1572

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCOTS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1196.6	76.1	6192	2 US-08-479-537A-1	Sequence 1, Appli
2	1196.6	76.1	6192	4 US-09-083-116-1	Sequence 1, Appli
3	1196.6	76.1	6192	4 US-09-134-916A-1	Sequence 1, Appli
4	625.2	39.8	6449	2 US-08-479-537A-4	Sequence 4, Appli
5	625.2	39.8	6449	4 US-09-083-116-4	Sequence 4, Appli
6	625.2	39.8	6449	4 US-09-134-916A-4	Sequence 4, Appli
7	77.2	5.0	320	4 US-09-165-264-12	Sequence 13, Appli
8	77.2	4.9	318	4 US-09-165-264-13	Sequence 13, Appli
9	77.2	4.9	320	4 US-09-165-264-14	Sequence 13, Appli
10	76.6	4.9	320	4 US-09-165-264-14	Sequence 7, Appli
11	76.6	4.8	319	4 US-09-165-264-11	Sequence 8, Appli
12	75.8	4.8	320	4 US-09-128-155-16	Sequence 16, Appli
13	73.8	4.7	152331	3 US-09-103-840A-2	Sequence 2, Appli
14	70.4	4.5	4403765	4 US-09-103-840A-2	Sequence 11, Appli
15	69.8	4.4	12001	1 US-08-458-568A-11	Sequence 12, Appli
16	67.8	4.3	4403765	4 US-09-103-840A-2	Sequence 12, Appli
17	67.7	4.3	7218	1 US-08-232-463-14	Sequence 14, Appli
18	65.8	4.2	1505	1 US-07-915-246-1	Sequence 1, Appli
19	62.6	4.0	494	4 US-09-056-556-176	Sequence 176, App
20	62.6	4.0	494	4 US-09-072-596-171	Sequence 171, App
21	62.6	4.0	4411529	4 US-09-103-840A-1	Sequence 1, Appli
22	61.2	3.9	985	4 US-09-056-556-182	Sequence 182, App
23	61.2	3.9	985	4 US-09-072-596-177	Sequence 177, App
24	61.2	3.9	1926	4 US-09-249-585A-4	Sequence 4, Appli
25	61	3.9	1931	2 US-09-130-114-2	Sequence 2, Appli
26	60.2	3.8	530	3 US-08-758-662-4	Sequence 4, Appli
27	60.2	3.8	4524	2 US-08-845-998-7	Sequence 7, Appli

C 28	60.2	3.8	4524	3 US-09-206-537-7	Sequence 7, Appli
C 29	60.2	3.8	4524	4 US-09-430-854-7	Sequence 7, Appli
C 30	59	3.8	4411529	4 US-09-103-840A-1	Sequence 1, Appli
C 31	58.4	3.7	6192	2 US-08-479-537A-1	Sequence 1, Appli
C 32	58.4	3.7	6192	4 US-09-083-116-1	Sequence 1, Appli
C 33	58.4	3.7	6192	4 US-09-134-916A-1	Sequence 1, Appli
C 34	58.4	3.7	6192	4 US-09-134-916A-1	Sequence 4, Appli
C 35	58.4	3.7	6449	2 US-08-479-537A-4	Sequence 4, Appli
C 36	58.4	3.7	6449	4 US-09-083-116-4	Sequence 4, Appli
C 37	58.4	3.7	6449	4 US-09-134-916A-4	Sequence 3, Appli
C 38	58	3.7	33529	4 US-09-144-085-3	Sequence 201, App
C 39	58	3.7	2367	4 US-09-056-556-201	Sequence 196, App
C 40	57.4	3.7	2166	2 US-08-878-546-9	Sequence 9, Appli
C 41	57.4	3.7	6530	2 US-08-146-930-1	Sequence 1, Appli
C 42	57.4	3.7	6530	3 US-08-458-240-1	Sequence 1, Appli
C 43	57.4	3.7	6530	3 PCT-US93-03993-1	Sequence 1, Appli
C 44	57.4	3.7	13967	2 US-08-804-227C-13	Sequence 13, Appli
C 45	57.4	3.7	44377	2 US-08-804-227C-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-479-537A-1
Sequence 1, Application US/08479537A
Patent No. 5861381/
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVET, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: BURNS, DOANE, SWECKER & MATSIS, L.L.P.
STREET: P. O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22113-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6192 base pairs

Query Match	Similarity	92.1%	Score 1196.6	DB 2	Length 6192
Best Local	Similarity	76.2%	Pred. No. 7e-267		
Matches 1223	Conservative	0	Mismatches 104	Indels	Gaps
QY 246	CCACAGCCCCGGTTACAGGCTCTCCACACTCAGGGAAGAGATGTACACTTGGACCCCGGC	305			
Db 4866	CNNNAGCCGANNCCGGGCTCCACCGCCCCCANNCCACAGTGTACACTCGGCCCGGA	4925			
QY 306	CACGGAACCACTTCAAGTTCACTGTACCTGAGGAGACAGATGTACACTTGGTCCCACT	365			
Db 4926	CNNNAGCCGANNCCGGGCTCCACCGCCCCCANNCCACAGGATGTACACTCGGCCCGGA	4985			
QY 366	CACACAGCCAGCCCTTGGGCTTCACAACCCCGCAGCCCAAGATGTCACTCAGCCCCGGA	425			
Db 4986	CNNNAGCCGANNCCGGGCTTCACAACCGCCCCCANNCCACAGGATGTCACTCGGCCCGGA	5045			
QY 426	CACACAGCCAGCCCGCGGCTTCACAACCGCCCCCGGACCAAGATGTCACTCGGCCCGGA	485			
Db 5046	CNNNAGCCGANNCCGGGCTTCACAACCGCCCCCANNCCACAGGATGTCACTCGGCCCGGA	5105			
QY 486	CACACAGCCAGCCCGCGGCTTCACAACCGCCCCCGCAGCGATGTCACTCGGCCCGGA	545			
Db 5106	CNNNAGCCGANNCCGGGCTTCACAACCGCCCCCANNCCACAGGATGTCACTCGGCCCGGA	5165			
QY 546	CACACAGCCAGCCCGCGGCTTCACAACCGCCCCCGGACCAAGATGTCACTCGGCCCGGA	605			
Db 5166	CNNNAGCCGANNCCGGGCTTCACAACCGCCCCCANNCCACAGGATGTCACTCGGCCCGGA	5225			
QY 606	CACACAGCCAGCCCTTGGGCTTCACAACCGCCCCCGCAGTGTCACTCGGCCCGGA	665			
Db 5226	CNNNAGCCGANNCCGGGCTTCACAACCGCCCCCGCAGTGTCACTCGGCCCGGA	5285			
QY 666	CTCTGCATCAGGCTCAGCTTCTACTCTGTGTCACAACGACCTCTGCAAGGCTTACAC	725			
Db 5286	CTCTGCATCAGGCTCAGCTTCTACTCTGTGTCACAACGACCTCTGCAAGGCTTACAC	5345			

Qy	726	AACCCAGCAGCAGAAAGACATCTCACCGAGATTCGACGACACACTCTGAATCTCTAC	785
Db	5346	AACCCAGCAGCAGAAAGACATCTCACCGAGATTCGACGACACACTCTGAATCTCTAC	5405
Qy	786	CACCCCTTGACAGCAGATACACAAAGACTAGATGCCAGTAGACATCACTATACACGGTACC	845
Db	5406	CACCCCTTGACAGCAGATACACAAAGACTAGATGCCAGTAGACATCACTATACACGGTACC	5465
Qy	846	TCTCTCTACCTCTCCCATATCAAGACATTTCTCCCAATGTGTACTGCGGGTCTCTTCTT	905
Db	5466	TCTCTCTACCTCTCCCATATCAAGACATTTCTCCCAATGTGTACTGCGGGTCTCTTCTT	5523
Qy	906	TTTCTCTGCTTTTTCACATTTTCAAACTCCAGTTTAATTTCTCTTGGAAGATCCGACAC	965
Db	5526	TTTCTCTGCTTTTTCACATTTTCAAACTCCAGTTTAATTTCTCTTGGAAGATCCGACAC	5585
Qy	966	CGACTACTACACAGAGCTGCAGAGACATTTCTGAATGTATTGTCAGATTATTAACA	1025
Db	5586	CGACTACTACACAGAGCTGCAGAGACATTTCTGAATGTATTGTCAGATTATTAACA	5645
Qy	1026	AGGGGGTTTCTTGCGGCTCTCCAAATATTAAGTTCAGGCGCAGATCTGCTGGGTGTACAAT	1089
Db	5646	AGGGGGTTTCTTGCGGCTCTCCAAATATTAAGTTCAGGCGCAGATCTGCTGGGTGTACAAT	5705
Qy	1086	GACTCTGGCTTTCGAGAAAGTACATCAATGTCCAAGACGTGGAGACACAGTTCAATCA	1145
Db	5706	GACTCTGGCTTTCGAGAAAGTACATCAATGTCCAAGACGTGGAGACACAGTTCAATCA	5765
Qy	1146	GTAATAAACGGAAGACGCTCTCGATATTAACCTGATACAGATCTCAGACGTCAAGCTTGAGTCA	1205
Db	5766	GTAATAAACGGAAGACGCTCTCGATATTAACCTGATACAGATCTCAGACGTCAAGCTTGAGTCA	5825
Qy	1206	TGTGCAATTTCTTTCTCTGCGCAAGTCTGGGGCTGGGGTCCAGAGCTGGGGCATCCGCT	1265
Db	5826	TGTGCAATTTCTTTCTCTGCGCCAGTCTGGGGCTGGGGTCCAGAGCTGGGGCATCCGCGCT	5885
Qy	1266	GCTGGTCTGTCTGTGTCTGTGTCTGTGTGTGTGGTGGCTGTGTCTATATCTATTTGCTGGCTGT	1325
Db	5886	GCTGGTCTGTGTGTGTCTGTGTCTGTGTGTGTGGTGGCTGTGTCTATATCTATTTGCTGGCTGT	5945
Qy	1326	CTGTCAGTGCTCCGAGAAAGACTAGCGGACGCTGTGACATTTTTCAGCCCGGGATACCTA	1385
Db	5946	CTGTCAGTGCTCCGAGAAAGACTAGCGGACGCTGTGACATTTTTCAGCCCGGGATACCTA	6005
Qy	1386	CCATCTTATGAGCGAGTACCCCATCTACACACCCATGGGCGCTATGTGCCCCCTAGCAG	1445
Db	6006	CCATCTTATGAGCGAGTACCCCATCTACACACCCATGGGCGCTATGTGCCCCCTAGCAG	6065
Qy	1446	TACGATCTGAGCCCTCTAGGAAGTTTCTGACAGTAATAGGTGCACAGCCTCTCTTA	1505
Db	6066	TACGATCTGAGCCCTCTAGGAAGTTTCTGACAGTAATAGGTGCACAGCCTCTCTTA	6125
Qy	1506	CACAAACCGACAGGTGAGAGCACTTCTGCACTGTATAGGCGACAGTGTGCCCTGTAGAC	1565
Db	6126	CACAAACCGACAGGTGAGAGCACTTCTGCACTGTATAGGCGACAGTGTGCCCTGTAGAC	6185
Qy	1566	TGAGTGG 1572	
Db	6186	TGAGTGG 6192	
RESULT 2			
US-09-083-116-1			
: Sequence 1, Application US/09083116			
: Patent No. 6203795			
: GENERAL INFORMATION:			
: APPLICANT: CHAMON, Pierre			
: APPLICANT: KIENY, Marie-Paule			
: APPLICANT: LATHER, Richard			
: APPLICANT: HAREVENT, Wera			
: TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE			
: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR			
: NUMBER OF SEQUENCES: 5			

CORRESPONDENCE ADDRESS:
 ADDRESSEE: BURNS, DOANE, SWECKER & MATTHIS, L.L.P.
 STREET: P.O. Box 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/083,116
 FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/479,537
 FILING DATE:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/FR91/00835
 FILING DATE: 23-OCT-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/039,320
 FILING DATE: 04-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/403,576
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Teskin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6192 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: sig_peptide
 LOCATION: 58..120

FEATURE:
 NAME/KEY: repeat_region
 LOCATION: 439..5239

OTHER INFORMATION: /note= "The nucleotides spanning 439-5239 constitute a repeated region wherein the repeat is 6 nucleotides and encodes 20 amino acids, 17 of which are fixed

OTHER INFORMATION: The number of such repeats varies from 1 to 80."

FEATURE:
 NAME/KEY: mat_deptide
 LOCATION: 121..6166

FEATURE:
 NAME/KEY: repeat_region
 LOCATION: 457

OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA

OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:
 NAME/KEY: repeat_region
 LOCATION: 487

OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN

OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG

OTHER INFORMATION: or ACG; and Asn = AAT or AAC."

FEATURE:
 NAME/KEY: repeat_region
 LOCATION: 496

OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA

OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

US-09-083-116-1

Query Match 76.1%; Score 1196.6; DB 4; Length 6192;
 Best Local Similarity 92.2%; Pred. No. 7e-267;
 Matches 1223; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 246 CCACAGCCCGGCTTCAGGCTCTCTCCACCACTCAGAGACAGAGATGTCACTTCGCCCCCGGC 305
 DB 4866 CANNAGGCCGANNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTTCGCCCCCGGA 4925

QY 306 CACGGAACCGAGCTTCAGGTTGAGTGCACCTGGGGACAGAGATGACCTCGAGTCCAGT 365
 DB 4926 CANNAGGCCGANNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTTCGCCCCCGGA 4985

QY 366 CACCAAGCCAGCCCTGGGCTCCACCACTTCGAGCCAGCCAGATGTCACTTCGCCCCCGGA 425
 DB 4986 CANNAGGCCGANNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTTCGCCCCCGGA 5045

QY 426 CAACAGCCAGCCCGGCTCCACCGCCCCCAGGACGAGTCACTTCGCCCCCGGA 485
 DB 5046 CANNAGGCCGANNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTTCGCCCCCGGA 5105

QY 486 CACCAAGCCCGCCCGGCTCCACCGCCCCCAGGACGAGTCACTTCGCCCCCGGA 545
 DB 5106 CANNAGGCCGANNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTTCGCCCCCGGA 5165

QY 546 CACCAAGCCCGCCCGGCTCCACCGCCCCCAGGACGAGTCACTTCGCCCCCGGA 605
 DB 5166 CANNAGGCCGANNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTTCGCCCCCGGA 5225

QY 606 CACCAAGCCCGCCCGGCTCCACCGCCCCCAGGACGAGTCACTTCGCCCCCGGA 665
 DB 5226 CANNAGGCCGANNNTGGGCTCCACCGCCCCCAGTCCACAGTCACTTCGCCCCCGGA 5285

QY 666 CTCTGCATCAGGCTCAGCTTCTACTGTGTCACAAAGGACCTTCGCCCCCGGA 725
 DB 5286 CTCTGCATCAGGCTCAGCTTCTACTGTGTCACAAAGGACCTTCGCCCCCGGA 5345

QY 726 AACCCAGCCAGCAAGACCTCCACCGGACATTCGAGCCAGCACTTCGAGTCACTTCGCCCCCGGA 785
 DB 5346 AACCCAGCCAGCAAGACCTCCACCGGACATTCGAGCCAGCACTTCGAGTCACTTCGCCCCCGGA 5405

QY 786 AACCCAGCCAGCCAGTACGACCAAGAGTATGACAGTCACTTCGAGTCACTTCGCCCCCGGA 845
 DB 5406 AACCCAGCCAGCCAGTACGACCAAGAGTATGACAGTCACTTCGAGTCACTTCGCCCCCGGA 5465

QY 846 TCCTCTACCTCTCCAAATCAAGCACTTCTCCAGTTGTTACTGGGCTCTTTCTT 905
 DB 5466 TCCTCTACCTCTCTCAATCAAGCACTTCTCCAGTTGTTACTGGGCTCTTTCTT 5525

QY 906 TTTCCGCTTTTCAATTCAATTCCTCGATTAACTCTCTGGAAGATCCAGCAC 965
 DB 5526 TTTCCGCTTTTCAATTCAATTCCTCGATTAACTCTCTGGAAGATCCAGCAC 5585

QY 966 CGACTACCTACCAAGCTGACAGAGACATTTTGAAATGTTTTCGAGATTATTAACA 1025
 DB 5586 CGACTACCTACCAAGCTGACAGAGACATTTTGAAATGTTTTCGAGATTATTAACA 5645

QY 1026 AGGGGGTTTTTCGGGCTCTCCAAATATTAAGTTCAGGACGAGATCTGTGTGTAATT 1085
 DB 5646 AGGGGGTTTTTCGGGCTCTCCAAATATTAAGTTCAGGACGAGATCTGTGTGTAATT 5705

QY 1086 GACTTCGGCTTCCAGAAAGTATCAATCAATGTTCCAGACGTGAGACAGATTCAATCA 1145
 DB 5706 GACTTCGGCTTCCAGAAAGTATCAATCAATGTTCCAGACGTGAGACAGATTCAATCA 5765

QY 1146 GTATTAACGGAAGAGAGCTCTCGATATACTGACGATCTGACAGCTCAAGCTGAGTCA 1205
 DB 5766 GTATTAACGGAAGAGAGCTCTCGATATACTGACGATCTGACAGCTCAAGCTGAGTCA 5825

QY 1206 TGTGCAATTTCTTTCTCTGCCCAGCTGGGGCTGGGGTCCAGAGCTGGGGCATTCGGGCT 1265
 DB 5826 TGTGCAATTTCTTTCTCTGCCCAGCTGGGGCTGGGGTCCAGAGCTGGGGCATTCGGGCT 5885

666 CTCTGATCAGGCTGAGCTTCTACTGTGAGCAAGGAGGAGCTCTGAGGAGCTACAC 725
5286 CTCTGATCAGGCTGAGCTTCTACTGTGAGCAAGGAGGAGCTCTGAGGAGCTACAC 5345
726 AACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 785
5346 AACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5405
786 CACCTTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 845
5406 CACCTTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5465
846 TCCTCTCAGCT 905
5466 TCCTCTCAGCT 5525
906 TTTCT 965
5526 TTTCT 5585
966 CGACTACTACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1025
5586 CGACTACTACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5645
1026 AGGGGGTTTTCTGGGGCTCTGCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1085
5646 AGGGGGTTTTCTGGGGCTCTGCAATTAATTAATTAATTAATTAATTAATTAATTAAT 5705
1086 GACTCTGGGCTCTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1145
5706 GACTCTGGGCTCTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5765
1146 GTATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1205
5766 GTATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5825
1206 TGTGCAATTTCT 1265
5826 TGTGCAATTTCT 5885
1266 GCTGAGTCTGCTGT 1325
5886 GCTGAGTCTGCTGT 5945
1326 CTGTGAGTCTGCTGT 1385
5946 CTGTGAGTCTGCTGT 6005
1386 CCATCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1445
6006 CCATCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6065
1446 TACGATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1505
6066 TACGATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6125
1506 CACGAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1565
6126 CACGAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6185
1566 TGAAGTG 1572
6186 TGAAGTG 6192

RESULT 4
US-08-479-537A-4
: Sequence 4, Application US/08479537A
: Patent No. 5861381
: GENERAL INFORMATION:
: APPLICANT: CHAMBERON, Pierre
: APPLICANT: KIENY, Marie-Paule

APPLICANT: LATHÉ, Richard
APPLICANT: HAREUVEN, Maria
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSES: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479, 537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6449 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note="The nucleotides spanning
439-5239 constitute a repeated region wherein the repeat is 6
nucleotides and encodes 20 amino acids, 17 of which are fixed
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..5661
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note="Nucleotide 457 is X1 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 487
OTHER INFORMATION: /note="Nucleotide 487 is Y = NNN
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACI
OTHER INFORMATION: or ACG, and Asn = AAT or AAC."

Thu May 8 18:53:07 2003

us-09-658-621b-1.rni

Page 6

```

: NAME/KEY: repeat_region
: LOCATION: 496
: OTHER INFORMATION: /note="Nucleotide 496 is X2 = NNN
: OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
: OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
:
: US-08-479-537A-4

```

Query Match	39.8%;	Score 625.2;	DB 2;	Length 6449;
Best Local Similarity	92.0%;	Pred. No. 3.8e-135;		
Matched CDS	0	Mismatched CDS	0	Coverage 0

	matches	83%	conservative	0%	misMatches	53%	Indels	0%	Gaps	0%
QY	1	GAATTC	CCCTGGTGTGTGAA	CTGTCTG	CCCCCTCC	CCACCAATTT	CAACAACAATG	60		
Db	1	GAATTC	CCCTGGTGTGTGAA	CTGTCTG	CCCCCTCC	CCACCAATTT	CAACAACAATG	60		
QY	61	ACAC	CGGGACCCAGTCTCTCTTCTTCTCTG	CTGCTCTCA	CAGTGTCA	AGTTCAGTGT	TT	120		
Db	61	ACAC	CGGGACCCAGTCTCTCTTCTTCTCTG	CTGCTCTCA	CAGTGTCA	AGTTCAGTGT	TT	120		
QY	121	ACAGGTT	CTGGCTATGC	CAAGCTCA	CCAGGTG	GAAGAAAGAA	GGAATTC	CGGTCCAG	180	
Db	121	ACAGGTT	CTGGCTATGC	CAAGCTCA	CCAGGTG	GAAGAAAGAA	GGAATTC	CGGTCCAG	180	
QY	181	ACAA	GTTCAGTCCCAAGCTCTCA	CTGAAAGAA	TGTTGTG	ATGAC	CACAGCGTACTC	240		
Db	181	AGAA	GTTCAGTCCCAAGCTCTCA	CTGAAAGAA	TGTTGTG	ATGAC	CACAGCGTACTC	240		
QY	241	TC	CAGCACA	CGCCCGGTTCA	GGCTCTCTCA	CACTCA	CGGACA	CGAATTCATCTG	300	
Db	241	TC	CAGCACA	CGCCCGGTTCA	GGCTCTCTCA	CACTCA	CGGACA	CGAATTCATCTG	300	
QY	301	CC	GGACACG	GAACGAGTTCAG	GTTCAGTGC	GCACCTT	GGGAGACAGATTCATCTG	360		
Db	301	CC	GGACACG	GAACGAGTTCAG	GTTCAGTGC	GCACCTT	GGGAGACAGATTCATCTG	360		
QY	361	CC	AGTCA	CCAGGCGCAGGCTCTG	GGCTTCCACA	CCCGGACGAGCCCA	CAATGTCAACTC	420		
Db	361	CC	AGTCA	CCAGGCGCAGGCTCTG	GGCTTCCACA	CCCGGACGAGCCCA	CAATGTCAACTC	420		
QY	421	CC	GGACAC	CAAGCTCA	GGCTCTG	CAACCGGCCCCCG	CGTTCACATCTG	480		
Db	421	CC	GGACAC	CAAGCTCA	GGCTCTG	CAACCGGCCCCCG	CGTTCACATCTG	480		
QY	481	CC	GGACAC	CAAGCTCA	GGCTCTG	CAACCGGCCCCCG	CGTTCACATCTG	540		
Db	481	CC	GGACAC	CAAGCTCA	GGCTCTG	CAACCGGCCCCCG	CGTTCACATCTG	540		
QY	541	CC	GGACAC	CAAGCTCA	GGCTCTG	CAACCGGCCCCCG	CGTTCACATCTG	600		
Db	541	CC	GGACAC	CAAGCTCA	GGCTCTG	CAACCGGCCCCCG	CGTTCACATCTG	600		
QY	601	CC	GGACAC	CAAGCTCA	GGCTCTG	CAACCGGCCCCCG	CGTTCACATCTG	660		
Db	601	CC	GGACAC	CAAGCTCA	GGCTCTG	CAACCGGCCCCCG	CGTTCACATCTG	660		
QY	661	TC	AGGCTCTG	CATCAG	GGCTCA	GCTTCTAC		689		
Db	661	CC	GGACAC	CAAGCTCA	GGCTCTG	CAACCGGCCCCCG	CGTTCACATCTG	660		

RESULT 5
US-09-083-116-4
Sequence 4, Application US/09083116
Patent No.: 6203795
GENERAL INFORMATION:
APPLICANT: CHAMRON, Pierre
APPLICANT: KIENEY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVENT, Maya
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

```

1 ADDRESSER:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
2 STREET:  P.O. Box 1404
3 CITY:  Alexandria
4 STATE:  Virginia
5 COUNTRY:  United States
6 ZIP:  22313-1404
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE:  Floppy disk
9 COMPUTER:  IBM PC compatible
10 OPERATING SYSTEM:  PC-DOS/MS-DOS
11 SOFTWARE:  Patentin Release #1.0, Version #1.30
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER:  US/09/083,116
14 FILING DATE:
15 CLASSIFICATION:
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER:  08/479,537
18 FILING DATE:
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER:  NO PCT/FR91/00835
21 FILING DATE:  23-OCT-1991
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER:  US 08/039,320
24 FILING DATE:  04-APR-1993
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER:  US 08/403,576
27 FILING DATE:  14-MAR-1995
28 ATTORNEY/AGENT INFORMATION:
29 NAME:  Teskin, Robin L.
30 REGISTRATION NUMBER:  35,030
31 REFERENCE/DOCKET NUMBER:  017753-025
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE:  (703) 836-6620
34 TELEFAX:  (703) 836-2021
35 INFORMATION FOR SEQ ID NO: 4:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 6449 base pairs
38 TYPE:  nucleic acid
39 STRANDEDNESS:  single
40 TOPOLOGY:  linear
41 MOLECULE TYPE:  DNA (genomic)
42 FEATURE:
43 NAME/KEY:  sig_peptide
44 LOCATION:  58..120
45 FEATURE:
46 NAME/KEY:  repeat_region
47 LOCATION:  439..5239
48 OTHER INFORMATION:  /note= "The nucleotides spanning
49 439-5239 constitute a repeated region wherein the repeat is 6
50 OTHER INFORMATION:  nucleotides and encodes 20 amino acids, 17 of which are fixed
51 OTHER INFORMATION:  The number of such repeats varies from 1 to 80."
52 FEATURE:
53 NAME/KEY:  mat_peptide
54 LOCATION:  121..5661
55 FEATURE:
56 NAME/KEY:  repeat_region
57 LOCATION:  457
58 OTHER INFORMATION:  /note= "Nucleotide 457 is X1 = NNN
59 OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
60 OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
61 FEATURE:
62 NAME/KEY:  repeat_region
63 LOCATION:  487
64 OTHER INFORMATION:  /note= "Nucleotide 487 is Y = NNN
65 OTHER INFORMATION:  which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
66 OTHER INFORMATION:  or ACG; and Asn = AAT or AAC."
67 FEATURE:
68 NAME/KEY:  repeat_region
69 LOCATION:  496
70 OTHER INFORMATION:  /note= "Nucleotide 496 is X2 = NNN
71 OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
72 OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
```



```
OY 1 GAATTCCTGCTGCTTGAATCTGTTCTGCCCCCTCCACCATTTCACCAACCAATG 60
Db 1 GAATTCCTGCTGCTTGAATCTGTTCTGCCCCCTCCACCATTTCACCAACCAATG 60
OY 61 AACACGGGACCCAGTCTCCCTTCTTCCTGCTGCTGCTCTCAAGTGTCTTAAGTGT 120
Db 61 AACACGGGACCCAGTCTCCCTTCTTCCTGCTGCTGCTCTCAAGTGTCTTAAGTGT 120
OY 121 ACAAGTTCTGCTCATGCAAGCTCTACCCAGGTGAGAAAAGAGACTTGGCTACCCG 180
Db 121 ACAAGTTCTGCTCATGCAAGCTCTACCCAGGTGAGAAAAGAGACTTGGCTACCCG 180
OY 181 AGAAGTTCAAGTCCAGCTCTACTGAGAGAAATGCTGATGATGACAGAGAGTACT 240
Db 181 AGAAGTTCAAGTCCAGCTCTACTGAGAGAAATGCTGATGATGACAGAGAGTACT 240
OY 241 TCCAGCCACAGCCCGGTTCAAGCTCTCTCCACACTCAAGGAGACAGATGTACTGAGC 300
Db 241 TCCAGCCACAGCCCGGTTCAAGCTCTCTCTCCACACTCAAGGAGACAGATGTACTGAGC 300
OY 301 CCGGCCCAGGAAACAGCTTCAAGTTCAAGTCACTGAGGAGACAGATGTACTGAGC 360
Db 301 CCGGCCCAGGAAACAGCTTCAAGTTCAAGTCACTGAGGAGACAGATGTACTGAGC 360
OY 361 CCAATCAACAGGCTCCTGAGCTTCCACACCCGCGACCTCAAGATGTACTGAGC 420
Db 361 CCAATCAACAGGCTCCTGAGCTTCCACACCCGCGACCTCAAGATGTACTGAGC 420
OY 421 CCGGACAAACAGCCCGGAGCTTCCACAGCCCGCGGACCAAGTGTACTGAGC 480
Db 421 CCGGACAAACAGCCCGGAGCTTCCACAGCCCGCGGACCAAGTGTACTGAGC 480
OY 481 CCGGACAAACAGCCCGGAGCTTCCACAGCCCGCGGACCAAGTGTACTGAGC 540
Db 481 CCGGACAAACAGCCCGGAGCTTCCACAGCCCGCGGACCAAGTGTACTGAGC 540
OY 541 CCGGACAAACAGCCCGGAGCTTCCACAGCCCGCGGACCAAGTGTACTGAGC 600
Db 541 CCGGACAAACAGCCCGGAGCTTCCACAGCCCGCGGACCAAGTGTACTGAGC 600
OY 601 CCGGACAAACAGCCCGGAGCTTCCACAGCCCGCGGACCAAGTGTACTGAGC 660
Db 601 CCGGACAAACAGCCCGGAGCTTCCACAGCCCGCGGACCAAGTGTACTGAGC 660
OY 661 TCAGGCTGCTGCTCAGGCTCTTCTTAC 689
Db 661 TCAGGCTGCTGCTCAGGCTCTTCTTAC 689
```

```
RESULT 7
US-09-165-264-13/c
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thiruvaiyath
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13
```

```
Query/ Match 5.0%; Score 78; DB 4; Length 320;
Best Local Similarity 53.2%; Pred. No. 2,4e-09;
Matches 165; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
```

```
OY 353 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 412
Db 313 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 254
OY 413 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 472
Db 253 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 194
OY 473 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 532
Db 193 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 134
OY 533 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 592
Db 133 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 74
OY 593 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 652
Db 73 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 14
OY 653 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 712
Db 13 CTTGAGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCA 4
```

```
RESULT 8
US-09-165-264-12/c
; Sequence 12, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thiruvaiyath
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12
```

```
Query Match 4.9%; Score 77.2; DB 4; Length 318;
Best Local Similarity 53.3%; Pred. No. 3.7e-09;
Matches 163; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
```

```
OY 357 GGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCACTC 416
Db 315 GGTCCAGTCAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCACTC 256
OY 417 AGCCCGGACAAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCACTC 476
Db 255 AGCCCGGACAAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCACTC 196
OY 477 GGTCCCGGACAAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCACTC 536
Db 196 GGTCCCGGACAAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCACTC 136
OY 537 GGTCCCGGACAAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCACTC 596
Db 135 GGTCCCGGACAAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCACTC 76
OY 597 GGTCCCGGACAAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCACTC 656
Db 75 GGTCCCGGACAAACAGGACCCCTGAGCTTCACACCCCGCAGGCCAGATGTCACTC 16
OY 657 GGTCCCTC 662
Db 15 CCCCCC 10
```


RESULT 9

```

US-09-165-264-7/c
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

```

```

Query Match          4.9%; Score 77; DB 4; Length 320;
Best Local Similarity 53.9%; Pred. No. 4.1e-09;
Matches 158; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```

```

QY 370 AGGCGACCCCTGGCTCCACACCCCGCAGCCCGAGATGTCACCTCAGCCCGGAGAC 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 316 AGGCGATGGCTAGCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 AAGCGACCCCGGCTCCACCGCCCGCCCGGCGGCTGTCACCTCGGCGCGGACAC 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 256 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 490 AGGCGCGCCCGGCTCCACCGCCCGCCCGGCGGCTGTCACCTCGGCGCGGACAC 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 196 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 550 AGGCGCGCCCGGCTCCACCGCCCGCCCGGCGGCTGTCACCTCGGCGCGGACAC 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 610 AGGCGCGCCCTTGGGCTCCACCGCCCGCTCAGTCCAGATGTCACCTCGGCTC 662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 76 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 24
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 10

```

US-09-165-264-14/c
; Sequence 14, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 14
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-14

```

```

Query Match          4.9%; Score 76.6; DB 4; Length 320;
Best Local Similarity 53.5%; Pred. No. 5e-09;
Matches 160; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
QY 347 ATGTACCTCGGCTCCAGTCCAGGACCGCCCTGGGCTCCACACCCCGGAGCCAGC 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 308 ATGGACACACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 249
QY 407 ATGTACCTCAGCCCGGAGCAACAGCCGCTCCGAGCTCAGCCGCTCCGAGCCAGC 466
DB 248 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 189
QY 467 GTGTACCTCGGCGCGGAGACACAGGCGCCCGGAGCTCAGCCGCTCCGAGCCAGC 526
DB 188 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 129
QY 527 GTGTACCTCGGCGCGGAGACACAGGCGCCCGGAGCTCAGCCGCTCCGAGCCAGC 586
DB 128 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 69
QY 587 GTGTACCTCGGCGCGGAGACACAGGCGCCCGGAGCTCAGCCGCTCCGAGCTCAGC 645
DB 68 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 10

```

RESULT 11

```

US-09-165-264-8/c
; Sequence 8, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 8
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-8

```

```

Query Match          4.8%; Score 76; DB 4; Length 319;
Best Local Similarity 52.9%; Pred. No. 6.9e-09;
Matches 163; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

```

```

QY 338 GGGGACAGATGTACCTCGATCCAGTCCAGGACGAGCCCTTGGGCTCCACACCCGCG 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 312 GGGTGTGTGAGACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 398 CAGCCACGATGTACCTCAGCCCGGAGCAACAGCAGCCCGGCTCCAGCGCCCGCC 457
DB 252 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 458 CGGCGCAGGTGTACCTCGGCGCGGAGACAGGCGCCCGGAGCTCCAGCGCCCGCG 517
DB 192 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 518 CGGCGCAGGTGTACCTCGGCGCGGAGACAGGCGCCCGGAGCTCCAGCGCCCGCC 577
DB 132 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 578 CGGCGCAGGTGTACCTCGGCGCGGAGACAGGCGGAGCTTGGGCTCCAGCGCCCGCT 637
DB 72 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 13
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 638 CAGTCCAC 645
DB 12 CCCCCCCC 5

```

RESULT 12

```

US-09-165-264-11/c
; Sequence 11, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:

```


QY 482 CGACACCAAGGCGCCGCTCCAGCGCCCGCCGACGATGTGACCTCGGCCC 541
 Db 3936819 CCGGACACCGCCGACACCTTGACACCGGACCGCGCGCGCTGACCACT 3936760
 QY 542 CGACACCAAGGCGCCGCTCCAGCGCCCGCCGACGATGTGACCTCGGCCC 601
 Db 3936759 TGTCCGCTGCGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCG 3936700
 QY 602 CGACACCAAGGCGCCGCTTCAGCGCCCGCTTCAGTCACAGATGTGACCTCGGCCC 661
 Db 3936699 CCGCTGCGCGCGCGCGCTTCAGCGCGCGCGCGCGCGCGCGCGCTGAG 3936640
 QY 662 CAGGCTGTGATGAGGCTGAGCTTCTACTGTGTGACACAGGACCTTGCAGGAGCTA 721
 Db 3936639 GAGGCTGCGCGCGCGCGCTTCAGGTGCGCGGACCGCTTGGCGCGCGTTCACAGCGTG 3936580
 QY 722 CCACACACCCGACGAGCAGACGACCTCCACCGAGCATTCGAGCAGCACTGTGATCTC 781
 Db 3936579 CCGTTAGCGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936520
 QY 782 CTACCAACC 789
 Db 3936519 CCTTGACC 3936512

RESULT 15
 US-08-458-568A-11
 ; Sequence 11, Application US/08458568A
 ; Patent No. 5821339
 ; GENERAL INFORMATION:
 ; APPLICANT: Schaffer, Priscilla A.
 ; APPLICANT: Yen, Lily
 ; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
 ; TITLE OF INVENTION: Infections
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339r1s
 ; STREET: One Liberty Place, 46th floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/458,568A
 ; FILING DATE: 02-JUNE-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/065,146
 ; FILING DATE: 05-MAY-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Leary Ph.D., Kathryn R.
 ; REGISTRATION NUMBER: 36,317
 ; REFERENCE/DOCKET NUMBER: DPCI-0029
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 568-3439
 ; TELEFAX: (215) 568-3100
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 12001 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Herpes simplex virus

STRAIN: Herpes Simplex Virus Type 1
 US-08-458-568A-11

Query Match 4.4%; Score 69.8; DB 1; Length 12001;
 Best Local Similarity 49.1%; Pred. No. 4.9e-07;
 Matches 185; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 267 CTCACACACACAGGAGACAGATGTGACCTTGGCCCGGACAGGAAACAGATTGAGTTC 326
 Db 1862 CCCAGGCTCTCCCGGCGCCGAGCCCTTCCCGGCGCTCCCGGCTTCCCGGCGGCTC 1921
 QY 327 AGCTGCACTTCAAGGAGACAGATGTGACCTTGGTCCAGTCAACAGGCGCGCTGAGCTC 386
 Db 1922 GGGCATCTCTACCTCAGTGCAGCGCAATCTCAGGTGAGAGATCAAACTCTCCGAGGCGC 1981
 QY 387 CACGACCCGCGAGCGCAAGATGTGACCTTCAAGCGCCCGAGCAACAGCCAGCCGCGCTC 446
 Db 1982 CCGGAGACCAACAGCGCCCTTGGCGCCCTTCCGACCTTGGCGCCCTTGGCGCC 2041
 QY 447 CACCGCCCGCGGCGCGACGAGTGTGACCTTGGCGCCCGAGACACAGGCGCGCCCGAGCTC 506
 Db 2042 CTTCCGCGCGCTTGGCGCCCTTCCCGCGCCCTTGGCGCCCTTCCCGCGCTTCCCGCGC 2101
 QY 507 CACGCGCCCGCGCGCGCGAGTGTGACCTTGGCGCCCGAGCAACAGGCGCGCGCGCTC 566
 Db 2102 CTTGCGCCCTTCCCGCGCCCTTGGCGCCCTTCCCGCGCCCTTCCCGCGCCCTTCCCGCGC 2161
 QY 567 CACCGCGCCCGCGCGCGCGAGTGTGACCTTGGCGCCCGAGCAACAGGCGCGCGCTTGGAGCTC 626
 Db 2162 CTCCCGCGCGCTTGGCGCCCTTCCCGCGCCCTTCCCGCGCCCTTCCCGCGCCCTTCCCGCGC 2221
 QY 627 CACCGCGCGCTTCCAGTTC 643
 Db 2222 CTGCGCCCTTCCCGCGC 2238

Search completed: May 8, 2003, 17:49:00
 Job time : 3112 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: May 8, 2003, 14:07:27 ; Search time 4354 Seconds

(without alignments)
10507.490 Million cell updates/sec

Title: US-09-658-621B-1

Perfect score: 1572
Sequence: 1 gattccctgcgtcgttgaa.....tcgccctctgagctgagtg 1572Scoring table: IDENTITY NUC
Gap 10.0 , Gape 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank :
1: gb_ba : *
2: gb_htg : *
3: gb_in : *
4: gb_om : *
5: gb_ov : *
6: gb_pat : *
7: gb_ph : *
8: gb_pl : *
9: gb_pr : *
10: gb_ro : *
11: gb_sts : *
12: gb_sy : *
13: gb_un : *
14: gb_vi : *
15: em_ba : *
16: em_fun : *
17: em_hum : *
18: em_in : *
19: em_mu : *
20: em_om : *
21: em_ov : *
22: em_ov : *
23: em_pat : *
24: em_ph : *
25: em_pl : *
26: em_ro : *
27: em_sts : *
28: em_un : *
29: em_vi : *
30: em_htg_hum : *
31: em_htg_inv : *
32: em_htg_other : *
33: em_htg_mus : *
34: em_htg_pln : *
35: em_htg_rnd : *
36: em_htg_mam : *
37: em_htg_vrt : *
38: em_sy : *
39: em_htgo_hum : *
40: em_htgo_mus : *
41: em_htgo_other : *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1572	100.0	1572	6 AX093798	AX093798 Sequence
2	1481.2	94.2	1721	6 AX335860	AX335860 Sequence
3	1481.2	94.2	1721	6 AX440427	AX440427 Sequence
4	1481.2	94.2	1721	9 HSTRYMA	X52229 Human mRNA
5	1412.2	89.8	1804	6 AX335367	AX335367 Sequence
6	1412.2	84.5	1795	9 HUMWUCAB	J05581 Human polyM
7	1328.4	84.5	1452	6 GIBWUC1A	L41589 Hylobates 1
8	1313	83.5	1452	6 A32135	A32135 H.sapiens m
9	1269	80.7	1834	12 AF423030	AF423030 Synthetic
10	1268	80.7	1414	12 AF423031	AF423031 Synthetic
11	1196.6	76.1	6192	6 AR142537	AR142537 Sequence
12	1196.6	76.1	6192	6 AR142537	AR142537 Sequence
13	1190.2	75.7	4139	6 AX334859	AX334859 Sequence
14	1190.2	75.7	4139	6 AX335372	AX335372 Sequence
15	1190.2	75.7	4139	6 AX336712	AX336712 Sequence
16	1190.2	75.7	4139	6 AX409474	AX409474 Sequence
17	1190.2	75.7	4139	6 AX440481	AX440481 Sequence
18	1190.2	75.7	4139	9 HUMPANMU	J05582 Human pancr
19	998	63.5	1296	9 HUMEPIS1A2	M34089 Human episi
20	998	63.5	1296	9 HUMEPIS1B2	M34089 Human episi
21	993.4	63.2	1320	6 E08764	E08764 cDNA encodi
22	919.4	58.5	2238	9 HSEETA	X52228 Human mRNA
23	750	47.7	8181	6 AX406624	AX406624 Sequence
24	746.4	47.5	1709	9 HUMPEM	M61170 Human polym
25	745.2	47.4	178229	2 AL713399	AL713399 Homo sapi
26	745.2	47.4	178229	2 AL713399	AL713399 Homo sapi
27	742	47.2	3343	9 HITMETMAGA	L35093 Human secre
28	694	44.1	5660	9 GIBWUC1A01	L41624 Hylobates 1
29	670.2	42.6	859	9 HS060261	U60261 Human MUC-1
30	627	39.9	2068	4 BTA400824	AJ400824 Bos tauru
31	625.2	39.8	6449	6 AR030812	AR030812 Sequence
32	625.2	39.8	6449	6 AR142538	AR142538 Sequence
33	619.4	39.4	805	9 HS060260	U60260 Human MUC-1
34	607.4	38.6	768	9 HSMUC1	X80761 H.sapiens M
35	605.8	38.5	759	9 AF125525	AF125525 Homo sapi
36	600.4	38.2	2119	4 OC85787	U85787 Oryctolagus
37	586.2	37.3	2242	10 BC005441	BC005441 Mus muscu
38	584	37.2	2166	10 MOSWUC1A	M84683 Mus muscu
39	556.8	35.4	2370	10 MAU36918	U36918 Mesocricetu
40	531	33.8	2751	9 AF176947	AF176947 Macaca mu
41	495.4	31.5	730	9 HS060259	U60259 Human MUC-1
42	494.4	31.5	497	6 AX440341	AX440341 Sequence
43	482.6	30.7	491	9 HUMEPIS1B1	M32739 Human episi
44	475.6	30.3	572	9 HUMDF3A	M31823 Human brea
45	459	29.2	5660	6 AX347259	AX347259 Sequence

ALIGNMENTS

RESULT 1
LOCUS AX093798 1572 bp DNA linear PAT 30-MAR-2001
DEFINITION Sequence 1 from Patent WO0118035.
ACCESSION AX093798
VERSION AX093798.1 GI:13510057
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1572)
Taylor-Papadimitriou, J., Heukamp, L.C., Offringa, R., Melief, C.J.,
Acres, B. and Thomas, M.
Muc-1 derived peptides

JOURNAL Patent: NO 0118035-A 1 15-MAR-2001;
TRANSGENE S.A. (FR) ; IMPERIAL CANCER RESEARCH TECHNOLOGY LIMITED
(GB)

FEATURES
source Location/Qualifiers
1..1572
/organism="Homo sapiens"
/db_xref="taxon:9606"
58..1545
/note="unamed protein product"

CDS

/codon_start=1
/protein_id="CA35422.1"
/db_xref="GI:13510058"
/translation="MTPGQSEFLLLLIVLVTVGSGHASTRGKREKTSATQSS
VPSSTENKAVSMSTSVLSHSPGSGSDVTLAPATBPASGSAATWQDVTSP
VTRPALGSTPPADHVTAPDNKPAAGSTAPADPPGSAADPAAGVTS
APDRPAAGSTAPPAHVTAPDNKPAAGSTAPADPPGSAADPAAGVTS
ARATTPASSTPSPSI PSHHSDPTPTTLASHTKTDASTHSTVPLTSSNHTSPCL
STGVSFPPLSPHISNLOPNSSLPDSIDYVQELORDISEMPLOIKGQPRGLSNIRP
RPSGVVQTLAFREGTINVDVETQFQVKTBAASHYNLTLSBVSHPVPPPSAOS
GAGVPMGIALVLVLCVLAVALVILALACQCRKNYQGLDITPADVTHPMSRYP
TYTHGRVYPPSSIDRSPEKVSAGNGSSLSITNPAVATSNL"

BASE COUNT 312 a 567 c 364 g 329 t
ORIGIN

Query Match 100.0%; Score 1572; DB 6; Length 1572;
Best Local Similarity 100.0%; Pred. No. 4.9e-277;

Matches 1572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCCTGAGCTGTTGAATCTGTCTGCCCCCTCCACCATTTTCCACCAACCATG 60
Db 1 GAATTCCTGAGCTGTTGAATCTGTCTGCCCCCTCCACCATTTTCCACCAACCATG 60
QY 61 ACACCGGGGACCACTGCTCTTCTCTGCTGCTGCTCTCTCAAGGCTTACAGTTGTT 120
Db 61 ACACCGGGGACCACTGCTCTTCTCTGCTGCTGCTCTCTCAAGGCTTACAGTTGTT 120
QY 121 ACAGTTCTGTCTATGCAAGCTCTACCCAGGTGAGAAAAGAGACTTGGCTACCCAG 180
Db 121 ACAGTTCTGTCTATGCAAGCTCTACCCAGGTGAGAAAAGAGACTTGGCTACCCAG 180
QY 121 ACAGTTCTGTCTATGCAAGCTCTACCCAGGTGAGAAAAGAGACTTGGCTACCCAG 180
Db 121 ACAGTTCTGTCTATGCAAGCTCTACCCAGGTGAGAAAAGAGACTTGGCTACCCAG 180
QY 181 AGAAGTTCTAGTGCCTAGCTCTACTGAGAAGATCTGTGATGATGACAGAGGATCTC 240
Db 181 AGAAGTTCTAGTGCCTAGCTCTACTGAGAAGATCTGTGATGATGACAGAGGATCTC 240
QY 241 TCACGCTCAGGCTGCTTCAAGCTCTTCACTCAAGGATGATGATCTTGGCTC 300
Db 241 TCACGCTCAGGCTGCTTCAAGCTCTTCACTCAAGGATGATGATCTTGGCTC 300
QY 301 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 360
Db 301 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 360
QY 361 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 420
Db 361 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 420
QY 421 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 480
Db 421 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 480
QY 481 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 540
Db 481 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 540
QY 541 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 600
Db 541 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 600
QY 601 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 660
Db 601 CCGGACCAAGGACCAAGCTTCAAGGCTGCTGAGGAGACAGATGATGATCTTGGCTC 660
QY 661 TCAGGCTGCTGATCAGGCTTCACTGTGTCACAAGGACCTCTGCGAGGAGCT 720

Db 661 TCAGGCTGCTGATCAGGCTTCACTGTGTCACAAGGACCTCTGCGAGGAGCT 720
QY 721 ACACCAACCCCGACGACGACGACCTTCCACCGACGATTTCCAGCTCTGATTA 780
Db 721 ACACCAACCCCGACGACGACGACCTTCCACCGACGATTTCCAGCTCTGATTA 780
QY 781 CCGACCAACCCCGACGACGACCTTCCACCGACGATTTCCAGCTCTGATTA 840
Db 781 CCGACCAACCCCGACGACGACCTTCCACCGACGATTTCCAGCTCTGATTA 840
QY 841 GATCTCTCTGACCTCTCTCAATGACAGACTTCTCCAGTGTCTATCTGGGCTCT 900
Db 841 GATCTCTCTGACCTCTCTCAATGACAGACTTCTCCAGTGTCTATCTGGGCTCT 900
QY 901 TTTCTTTTCTGCTTTTCACTTCAAACTTCAGTTAATTCCTCTGGAAGATCC 960
Db 901 TTTCTTTTCTGCTTTTCACTTCAAACTTCAGTTAATTCCTCTGGAAGATCC 960
QY 961 AGCAGGACTTACCAAGAGCTGACAGAGACATTTTGAAATGTTTTGCAATTTAT 1020
Db 961 AGCAGGACTTACCAAGAGCTGACAGAGACATTTTGAAATGTTTTGCAATTTAT 1020
QY 1021 AACCAAGGGGATTTCTGGGCTCTCCATATTAATTTAGGCGCAGATCTGTGGTA 1080
Db 1021 AACCAAGGGGATTTCTGGGCTCTCCATATTAATTTAGGCGCAGATCTGTGGTA 1080
QY 1081 CAATTAATCTGGGCTTCCGAGAGGATCAATGATTCACAGCTGAGACACAGTTTC 1140
Db 1081 CAATTAATCTGGGCTTCCGAGAGGATCAATGATTCACAGCTGAGACACAGTTTC 1140
QY 1141 AATCAGATPAAACGAGAGACAGCTCTGATATTAATTCAGATCTCAGAGTCAAGG 1200
Db 1141 AATCAGATPAAACGAGAGACAGCTCTGATATTAATTCAGATCTCAGAGTCAAGG 1200
QY 1201 AGTCATGTGCAATTTCTTTCTCTGCCCAGTCTGGGCTGCGAGCTGGGGATC 1260
Db 1201 AGTCATGTGCAATTTCTTTCTCTGCCCAGTCTGGGCTGCGAGCTGGGGATC 1260
QY 1261 GCGCTGCTGGTGTGCTGTGTTCTGCTGCGCTGAGCTGATCTGATGCTGCTT 1320
Db 1261 GCGCTGCTGGTGTGCTGTGTTCTGCTGCGCTGAGCTGATCTGATGCTGCTT 1320
QY 1321 GCTGTCTGTGAGTGGCGGAGAAAGAACTACGGGACGTGACATCTTCCAGCCGGAT 1380
Db 1321 GCTGTCTGTGAGTGGCGGAGAAAGAACTACGGGACGTGACATCTTCCAGCCGGAT 1380
QY 1381 ACTTACCATCTATGAGGAGATCCCACTACACACCATGGGCGCTATGTCCTCCT 1440
Db 1381 ACTTACCATCTATGAGGAGATCCCACTACACACCATGGGCGCTATGTCCTCCT 1440
QY 1441 AGCAGTACGATGTGAGCCCTTATGAGAGGTTTCTGAGGTAATGATGAGAGGCTC 1500
Db 1441 AGCAGTACGATGTGAGCCCTTATGAGAGGTTTCTGAGGTAATGATGAGAGGCTC 1500
QY 1501 TCTTACCAAAACCCAGAGTGGCAGACATTTGCAACTTGTAGGGGACGTCGCCCTC 1560
Db 1501 TCTTACCAAAACCCAGAGTGGCAGACATTTGCAACTTGTAGGGGACGTCGCCCTC 1560
QY 1561 TGAAGTGAATGG 1572
Db 1561 TGAAGTGAATGG 1572

RESULT 2
AX335860
LOCUS AX335860 1721 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 6369 from Patent WO0194629.
ACCESSION AX335860
VERSION AX335860.1 GI:18126579
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1
 AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ehner, R., Endress, G.,
 Horvitz, S., Soppet, D. R. and Weaver, Z.
 TITLE Cancer gene determination and therapeutic screening using signature
 gene sets
 JOURNAL Patent: WO 0194629-A 6369 13-DEC-2001;
 Avalon Pharmaceuticals (US)
 FEATURES Location/Qualifiers
 source 1..1721
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 339 a 634 c 395 g 353 t
 ORIGIN

Query Match 94.2% Score 1481.2; DB 6; Length 1721;
 Best Local Similarity 95.5%; Pred. No. 1.8e-260;
 Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

QY 1 GAATTCCTGGCTGTTGAATGTTCTGCCCCCTCCCAACCATTTCCACCAACCATG 60
 DB 1 GAATTCCTGGCTGTTGAATGTTCTGCCCCCTCCCAACCATTTCCACCAACCATG 60
 QY 61 ACACCGGGACCCAGTCTCTCTTCTGCTGCTCTCTCAAGTGTTCAGATTGT 120
 DB 61 ACACCGGGACCCAGTCTCTCTTCTGCTGCTCTCTCAAGTGTTCAGATTGT 120
 QY 121 ACAGGTTCTGTCATGCAAGCTCTACCCAGGTGGAGAAAGAGACTTCGGCTTACCAG 180
 DB 121 ACAGGTTCTGTCATGCAAGCTCTACCCAGGTGGAGAAAGAGACTTCGGCTTACCAG 180
 QY 181 AGAAGTTCAGTGTCCAGCTCTACTGAGAAAGATGCTGTGATGACAGAGAGTATC 240
 DB 181 AGAAGTTCAGTGTCCAGCTCTACTGAGAAAGATGCTGTGATGACAGAGAGTATC 240
 QY 241 TCCAGCCACAGCCCGGGTTCAGGCTCTCCACCACTAGGAGAGAGATGTACTGTGGC 300
 DB 241 TCCAGCCACAGCCCGGGTTCAGGCTCTCCACCACTAGGAGAGAGATGTACTGTGGC 300
 QY 301 CCGGCGCCAGGAAACAGCTTTCAGGCTGCACTGTGGGAAACAGATGTCACTCGGTC 360
 DB 301 CCGGCGCCAGGAAACAGCTTTCAGGCTGCACTGTGGGAAACAGATGTCACTCGGTC 360
 QY 361 CCAATCACCAGGCGAGCTCTGGGCTCCACCAACCCGAGAGAGAGATGTACTGTGGC 420
 DB 361 CCAATCACCAGGCGAGCTCTGGGCTCCACCAACCCGAGAGAGATGTACTGTGGC 420
 QY 421 CCGGACCAAAAGCTCAGCTCCGAGGCTCCACCGCCCGGAGGATGTCACTCGGCTC 480
 DB 421 CCGGACCAAAAGCTCAGCTCCGAGGCTCCACCGCCCGGAGGATGTCACTCGGCTC 480
 QY 481 CCGGACCAAGGCGCGCTCCGAGGCTCCACCGCCCGGAGGATGTCACTCGGCTC 513
 DB 481 CCGGACCAAGGCGCGCTCCGAGGCTCCACCGCCCGGAGGATGTCACTCGGCTC 513
 QY 514 -----CCGCGCGCCACAGGTGTCACTCGGCTC 540
 DB 541 CCGGACCAAGGCGCGCTCCGAGGCTCCACCGCCCGGAGGATGTCACTCGGCTC 600
 QY 541 CCGGACCAAGGCGCGCTCCGAGGCTCCACCGCCCGGAGGATGTCACTCGGCTC 600
 DB 601 CCGGACCAAGGCGCGCTCCGAGGCTCCACCGCCCGGAGGATGTCACTCGGCTC 660
 QY 601 CCGGACCAAGGCGCGCTCCGAGGCTCCACCGCCCGGAGGATGTCACTCGGCTC 660
 DB 661 CCGGACCAAGGCGCGCTCCGAGGCTCCACCGCCCGGAGGATGTCACTCGGCTC 720
 QY 661 TGAAGCTCTGATCAGGCTCAGCTTCTACTGTGCAAGAGGCAAGTGTCACTCGGCTC 720
 DB 721 TGAAGCTCTGATCAGGCTCAGCTTCTACTGTGCAAGAGGCAAGTGTCACTCGGCTC 780
 QY 721 ACCACAAACCCAGCGAGAGAGATCCACCCAGATTCCTCAGCCACCACTGTACT 780

DB 781 ACCACAAACCCAGCGAGAGAGATCCACCCAGATTCCTCAGCCACCACTGTACT 840
 QY 781 CCTACCACTCTTCCAGCAATAGCAACCAAGCTGATGGCAGTACCTACCAATAGCAG 840
 DB 841 CCTACCACTCTTCCAGCAATAGCAACCAAGCTGATGGCAGTACCTACCAATAGCAG 900
 QY 841 GATACCTCTCAGCTCTCTCCATTCAGCAAGCTTCTCCGAGTGTCTACTGGGCTCT 900
 DB 901 GATACCTCTCAGCTCTCTCCATTCAGCAAGCTTCTCCGAGTGTCTACTGGGCTCT 960
 QY 901 TTTCTTTCTCTGCTTTTCAATTCAAACTCTCAATTAAATTCCTCTGGAAGATCCC 960
 DB 961 TTTCTTTCTCTGCTTTTCAATTCAAACTCTCAATTAAATTCCTCTGGAAGATCCC 1020
 QY 961 AGACCGAATCTACCAAGAGCTGAGAGAGATTTCTGAAGTTTGGCAGATTTAT 1020
 DB 1021 AGACCGAATCTACCAAGAGCTGAGAGAGATTTCTGAAGTTTGGCAGATTTAT 1080
 QY 1021 AAACAAAGGAGGTTTCTGGGCTCTCAATATTAGTTCAAGCCAGGATCTGGTGTGA 1080
 DB 1081 AAACAAAGGAGGTTTCTGGGCTCTCAATATTAGTTCAAGCCAGGATCTGGTGTGA 1140
 QY 1081 CAATTGACTCTGGCTTCCGAGAGGTACATCAATGTCACAGAGCTGAGACAGATTTC 1140
 DB 1141 CAATTGACTCTGGCTTCCGAGAGGTACATCAATGTCACAGAGCTGAGACAGATTTC 1200
 QY 1141 AATCAATTAACGAAAGAGCTCTCTGATTAACCTGACATCTCAGAGCTGACGCTG 1200
 DB 1201 AATCAATTAACGAAAGAGCTCTCTGATTAACCTGACATCTCAGAGCTGACGCTG 1260
 QY 1201 AGTCAATGTCATTTCTTTCTCTGCTGAGCTGGGAGCTGGGAGCTGGGAGATC 1260
 DB 1261 AGTCAATGTCATTTCTTTCTCTGCTGAGCTGGGAGCTGGGAGCTGGGAGATC 1320
 QY 1261 GCGCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
 DB 1321 GCGCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
 QY 1321 GCTGTCTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 DB 1381 GCTGTCTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
 QY 1381 ACTTACCAATCTTATGAGCAATACCCACTTACCAACCCAGTGGGCTGATGTGCCCT 1440
 DB 1441 ACTTACCAATCTTATGAGCAATACCCACTTACCAACCCAGTGGGCTGATGTGCCCT 1500
 QY 1441 AGCAGTACCAATGTAAGCTTATGAGAAAGTTTCTGAGATTAATGTGGAGAGCTTC 1500
 DB 1501 AGCAGTACCAATGTAAGCTTATGAGAAAGTTTCTGAGATTAATGTGGAGAGCTTC 1560
 QY 1501 TCTTACCAAAACCAAGATGAGCACTTCTGCAACTTTGTAAGGAGAGCTGCTC 1560
 DB 1561 TCTTACCAAAACCAAGATGAGCACTTCTGCAACTTTGTAAGGAGAGCTGCTC 1620
 QY 1561 TGAAGTGTGG 1572
 DB 1621 TGAAGTGTGG 1632

RESULT 3
 AX440427 1721 bp DNA linear PAT 28-JUN-2002
 LOCUS AX440427
 DEFINITION Sequence 280 from Patent WO0190154.
 ACCESSION AX440427
 VERSION AX440427.1 GI:21665237
 KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Xu, J., Mitcham, J. L., Harlocker, S. L., Dillon, D. C., Sechrist, H.,

Lodes,M.J., Algate,P.A., Fling,S.P., Mannion,J., Benson,D.R. and Carter,D.
Compositions and methods for the therapy and diagnosis of ovarian cancer

JOURNAL Patent: WO 0190154-A 280 29-NOV-2001;
CORIXA CORPORATION (US)

FEATURES Location/Qualifiers

Source 1..1721
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 339 a 634 c 395 g 353 t

ORIGIN

Query Match 94.2%; Score 1481.2; DB 6; Length 1721;
Best Local Similarity 95.5%; Pred. No. 1.8e-260;

Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

```

QY 1 GAATTCCTGGCTGCTGAAATCTGTGTGCGCCCTCCCAACCCATTTCACACACATG 60
DB 1 GAATTCCTGGCTGCTGAAATCTGTGTGCGCCCTCCCAACCCATTTCACACACATG 60
QY 61 AACCGGAGCAACCAAGTCTCTTCTCCGCTGCTGCTCCACAGTCTTACAGTTGTT 120
DB 61 AACCGGAGCAACCAAGTCTCTTCTCCGCTGCTGCTCCACAGTCTTACAGTTGTT 120
QY 121 ACGAGTTCTGTCTATGCAAGCTCTACCCAGGTGAGAAAAGAGACTTGGCTACG 180
DB 121 ACGAGTTCTGTCTATGCAAGCTCTACCCAGGTGAGAAAAGAGACTTGGCTACG 180
QY 181 AGAAGTTAGAGGCCAGCTCTACTAGAGAAATGCTGTGATATGACAGACAGCTACT 240
DB 181 AGAAGTTAGAGGCCAGCTCTACTAGAGAAATGCTGTGATATGACAGACAGCTACT 240
QY 241 TCACAGCAAGCCCGGTTTAAAGCTCTTCAACCACTCAGGAGACAGATGCTACTTGGCC 300
DB 241 TCACAGCAAGCCCGGTTTAAAGCTCTTCAACCACTCAGGAGACAGATGCTACTTGGCC 300
QY 301 CCGGAGCAAGCAAGCTTCAAGTTCAAGTGCACCTGGGAGACAGATGCTACTTGGCT 360
DB 301 CCGGAGCAAGCAAGCTTCAAGTTCAAGTGCACCTGGGAGACAGATGCTACTTGGCT 360
QY 361 CCAATGACCAAGGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 420
DB 361 CCAATGACCAAGGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 420
QY 421 CCGGAGCAAGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 480
DB 421 CCGGAGCAAGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 480
QY 481 CCGGAGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 540
DB 481 CCGGAGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 540
QY 514 -----CCGCGCGCCCAAGGTTTCACTTGGCC 540
DB 514 -----CCGCGCGCCCAAGGTTTCACTTGGCC 540
QY 541 CCGGAGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 600
DB 541 CCGGAGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 600
QY 601 CCGGAGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 660
DB 601 CCGGAGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 660
QY 661 CCGGAGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 720
DB 661 CCGGAGCAAGCTTGGGCTTCAACACCCCGCAAGCTTCAAGTCACTTGGCC 720
QY 721 TCAGGCTCTGATCAGGCTTCACTTGTGTGCAAGGAGCAAGCTTGGCCAGGCT 780
DB 721 TCAGGCTCTGATCAGGCTTCACTTGTGTGCAAGGAGCAAGCTTGGCCAGGCT 780
QY 781 ACCGACACCCGAGCAGGAGCAAGCTTCAAGCTTCAAGCTTGGCCAGCTTGGATCT 840
DB 781 ACCGACACCCGAGCAGGAGCAAGCTTCAAGCTTCAAGCTTGGCCAGCTTGGATCT 840

```

```

QY 781 CCTACCACTTGGCGAGCCATAGACCAAGCTGATGCAATGACCTACCATAGCAG 840
DB 841 CCTACCACTTGGCGAGCCATAGACCAAGCTGATGCAATGACCTACCATAGCAG 900
QY 841 GATCTCTCTCACTTCTTCAATGCAAGCACTTCTCCCAAGTTGCTTCAAGGCTCT 900
DB 901 GATCTCTCTCACTTCTTCAATGCAAGCACTTCTCCCAAGTTGCTTCAAGGCTCT 960
QY 901 TTTCTTTTCTGCTTCTTCTTCAATGCAAGCACTTCTCCCAAGTTGCTTCAAGGCT 960
DB 961 TTTCTTTTCTGCTTCTTCTTCAATGCAAGCACTTCTCCCAAGTTGCTTCAAGGCT 1020
QY 961 AGCAGCACTACTACCAAGGCTGCAAGAGCAATTTCTGAATGTTTTCAGATTTAT 1020
DB 1021 AGCAGCACTACTACCAAGGCTGCAAGAGCAATTTCTGAATGTTTTCAGATTTAT 1080
QY 1021 AAAAGAGGAGGTTTCTTGGGCTCTCAATATTAAGTTCAAGGCAAGTCTGTGTGTA 1080
DB 1081 AAAAGAGGAGGTTTCTTGGGCTCTCAATATTAAGTTCAAGGCAAGTCTGTGTGTA 1140
QY 1081 CAATGACTTGGGCTTCTGAGAAAGTACATGATTCACAGACGTCGAGACAGTTTC 1140
DB 1141 CAATGACTTGGGCTTCTGAGAAAGTACATGATTCACAGACGTCGAGACAGTTTC 1200
QY 1141 AATCAGTATTAAGGAGAGAGAGAGCTCTGATATTAAGTCAAGATCTCAGACGTTG 1200
DB 1201 AATCAGTATTAAGGAGAGAGAGAGCTCTGATATTAAGTCAAGATCTCAGACGTTG 1260
QY 1201 AGTCATGTCATTTCTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
DB 1261 AGTCATGTCATTTCTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
QY 1261 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
DB 1321 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
QY 1321 GCTGTCTGTCAGTCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1381 GCTGTCTGTCAGTCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1381 ACCTACATCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1441 ACCTACATCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1441 AGCAGTACCGATGTCAGGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1501 AGCAGTACCGATGTCAGGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1501 TCTTACCAAAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1561 TCTTACCAAAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1561 TGAAGTGAAGTG 1572
DB 1621 TGAAGTGAAGTG 1632

```

RESULT 4

HSTEMYA

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

HSTEMYA 1721 bp mRNA linear PRI 14-AUG-1995
Human mRNA for transmembrane epithelial tumor mucin antigen.

X52229.1 GI:37053
transmembrane protein; tumor antigen.

Homo sapiens.

Homo sapiens.

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 1721)

Wreschner D.H.

Direct Submission

Submitted (16-MAR-1990) Wreschner D.H., Tel Aviv University, Dept

of Microbiology, Faculty of Life Sciences, Tel Aviv 69978, Israel

Qy	1321	GGGCTGTGTGCAAGTCCGCCCGAAAGAACTACGGGCAAGCTGTGACATCTTTTCAGAGCCGGAGAT	1380
Db	1381	GGCTGTGTGCAAGTCCGCCCGAAAGAACTACGGGCAAGCTGTGACATCTTTTCAGAGCCGGAGAT	1440
Qy	1381	ACCTACCAATCCCTATGAGCGAGTACGCCCACTACACACACCAATGGGCGGCTAATGTGCCCCCT	1440
Db	1441	ACCTACCAATCCCTATGAGCGAGTACGCCCACTACACACACCAATGGGCGGCTAATGTGCCCCCT	1500
Qy	1441	AGCAGTACCCATTCGTACCCCTTATGGAAGGTTCTGACAGTAAATGTGTGGACGAGAGCTC	1500
Db	1501	AGCAGTACCCATTCGTACCCCTTATGGAAGGTTCTGACAGTAAATGTGTGGACGAGAGCTC	1560
Qy	1501	TCTTTACACAAACCAGACAGATGGAGAGCACTTCTGCAACTGTATGAGGGGACAGTGGCCCTC	1560
Db	1561	TCTTTACACAAACCAGACAGATGGAGAGCACTTCTGCAACTGTATGAGGGGACAGTGGCCCTC	1620
Qy	1561	TGAGCTGTAGTGG	1572
Db	1621	TGAGCTGTAGTGG	1632

RESULT 5	AX35367	1804 bp	DNA	linear	PAT 09-JAN-2002
LOCUS	AX35367	Sequence 5876	from Patent WO0194629.		
DEFINITION	AX35367				
ACCESSION	AX35367.1	GI:18126086			
VERSION					
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1	Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,			
AUTHORS	Harrison, S., Soppet, D.R. and Weaver, Z.				
TITLE	Cancer gene determination and therapeutic screening using signature gene sets				
JOURNAL	Patent: WO 0194629-A	5876	13-DEC-2001;		
	Avalon Pharmaceuticals (US)				
FEATURES	Location/Qualifiers				
source	1..1804				

BASE COUNT	361	a	622	c	434	g	387	t
ORIGIN	/db_xref=taxon:9606							
Query Match	89.8%	Score	1412.2	DB	6	Length	1804	
Best Local Similarity	95.3%	Pred. No.	76-448					
Matches 1490	Conservative	0	Mismatches	13	Indels	60	Gaps	1

Db	25	GCCGTGCTAAMCTGTTCTGCCCCCTCCACCCATTTCACACACACATGACACCCGGGC	84
OY	70	ACCAGCTCTCTTTCTTCTGCTGCTGCTCTCACAGTGCTTACAGTGTTCACAGTCT	129
Db	85	ACCAGCTCTCTTTCTTCTGCTGCTGCTCTCACAGTGCTTACAGTGTTCACAGTCT	144
OY	130	GGTCATGCAAGCTCTACCCCGAGTGAGAAAAAGAGACTTGCGCTGCCAGAGAACTCA	189
Db	145	GGTCATGCAAGCTCTACCCCGAGTGAGAAAAAGAGACTTGCGCTGCCAGAGAACTCA	204
OY	190	GTGCCCAGCTCTACTGAGAGAAATGCTGTGATGTATGACACAGACGCTACTCTTCACGAC	249
Db	205	GTGCCCAGCTCTACTGAGAGAAATGCTGTGATGTATGACACAGACGCTACTCTTCACGAC	264
OY	250	AGCCCCGGTTCAAGGCTCTCCACACCTCAGGGGACAGATATTCATCTGGGCCCGGACAG	309
Db	265	AGCCCCGGTTCAAGGCTCTCTCCACACCTCAGGGGACAGATATTCATCTGGGCCCGGACAG	324
OY	310	GAACCAAGCTTCAGATTCAGCTGGCCACTCTGGGAGACAGATATTCATCTGGTCCCACTACC	369
Db	325	GAACCAAGCTTCAGATTCAGCTGGCCACTCTGGGAGACAGATATTCATCTGGTCCCACTACC	384

QY	QY	QY	QY
370	AAGCCAGCCCTGAGGCTCCACACCCCGCCAGCCCAAGATGTCACTTACGCCCCGAGCAAC	429	
385	AGGCGAGCCCTGAGGCTCCACACCCCGCCAGCCCAAGATGTCACTTACGCCCCGAGCAAC	444	
430	AAGCCAGCCCTGAGGCTCCACACCCCGCCAGCCCAAGATGTCACTTACGCCCCGAGCAAC	489	
445	AAGCCAGCCCTGAGGCTCCACACCCCGCCAGCCCAAGATGTCACTTACGCCCCGAGCAAC	494	

OY	550	AGGCGGCGCCCGGCGCTTCACCGCGCCCCCGGCGCAAGGTGCACCTTGCGCCCGGACAC	603
Db	505	AGGCGGCGCCCGGCGCTTCACCGCGCCCCCGGCGCAAGGTGCACCTTGCGCCCGGACAC	564
OY	610	AGGCGGCGCTTGCGGCTTCACCGCGCCCTTCAGTCCAAATGTCACTCGGCTCAGGCTCT	663
Db	565	AGGCGGCGCTTGCGGCTTCACCGCGCCCTTCAGTCCAAATGTCACTCGGCTCAGGCTCT	624
OY	670	GCATCAGGCTCAGCTTCTACTGTGTGCACAAAGGCACTTCTCCAGGCTTACACAAAC	723
Db	625	GCATCAGGCTCAGCTTCTACTGTGTGCACAAAGGCACTTCTCCAGGCTTACACAAAC	684
OY	730	CCAGCAGACAGAGACACTCACCCAGATTCCAGGCCACTGTGATCTCTTACACAC	783
Db	685	CCAGCAGACAGAGACACTCAATTCATTTCCAGCACACTGTGATCTCTTACACAC	744
OY	790	CTTTCGCACCTTAGACCAAGCTGATGACGAGAGACACTCAACATTAACAGGTAACCTCT	845
Db	745	CTTTCGCACCTTAGACCAAGCTGATGACGAGAGACACTCAACATTAACAGGTAACCTCT	806
OY	850	CTCAGCTCTCTCAATCAAGACACTTCTCCCAAGTGTCTACTGCGGCTCTCTTTCTTTTC	903
Db	805	CTCAGCTCTCTCAATCAAGACACTTCTCCCAAGTGTCTACTGCGGCTCTCTTTCTTTTC	864
OY	910	CTGTCTTTTCACATTTCAAAAGCTCCAGTTTAATTTCTCTGGAAGATCCACAGCCAG	965
Db	865	CTGTCTTTTCACATTTCAAAAGCTCCAGTTTAATTTCTCTCTGGAAGATCCACAGCCAG	922

Db	925	TACTACAAAGAGCTGCAGAGAAACATTTCTGAAATCTTTTTCAGATTTATTAACAAAGG	98
Qy	1030	GGTTTCTCGGAGCCTCTCCAAATATTAAAGTCAAGGCCAGATCTGTGTGTACATTAAGCT	106
Db	985	GGTTTTCGTGGGCCTCTCCAAATATTAAAGTCAAGGCCAGATCTGTGTGTACATTTAGCT	104
Qy	1090	CTGGCCTTCCAGAAAGTACATCAATGTCCAGAGTGTGGAGACACAGTTCAATCAATGAT	114
Db	1045	CTGGCCTTCCAGAAAGTACATCAATGTCCAGAGTGTGGAGACACAGTTCAATCAATGAT	110
Qy	1150	AAAACGAAAGAGCCTCTGCATATTAACCTGACATCTCAGAGTCAACGTCAGATCATGNG	120
Db	1105	AAAACGAAAGAGCCTCTGCATATTAACCTGACATCTCAGAGTCAACGTCAGATCATGNG	116
Qy	1210	CAATTTCTTTTCTCTGCCCAGTCTGGGCTGGGGTGCAGGCTGTGGGCATCCGGGCTGCTG	122
Db	1165	CAATTTCTTTTCTCTGCCCAGTCTGGGCTGGGGTGCAGGCTGTGGGCATCCGGGCTGCTG	122
Qy	1270	GATCGGTGTGTGTCTGATGTGGTGGGCTGGGCAATGTCTATCTCATTTGCTTGAGTGTGT	134
Db	1225	GATCGGTGTGTGTCTGATGTGGTGGGCTGGGCAATGTCTATCTCATTTGCTTGAGTGTGT	122
Qy	1330	CAGTGCAGCCGAAAAGACTACGAGGACGTGGACATCTTTTCCAGCCCGAGATCTTACAT	138
Db	1285	CAGTGCAGCCGAAAAGACTACGAGGACGTGGACATCTTTTCCAGCCCGAGATCTTACAT	134
Qy	1390	CTATATAGAGAGTACCCCACTTACACAGCCATGGGCGGTATGTGCCCTCCACAGATACC	144
Db	1345	CTATATAGAGAGTACCCCACTTACACAGCCATGGGCGGTATGTGCCCTCCACAGATACC	140

Oy	1450	GATCGTAGCCCTATGAGAAAGTTCTGAGGTATATGTGTGCAGACGCTCTCTTACACA	1509
Db	1405	GATCGTAGCCCTATGAGAAAGTTCTGAGGTATATGTGTGCAGACGCTCTCTTACACA	1464
Oy	1510	AATCCAGCAGTGTGCAGCCACTTGTGCACATTGTATGGGGCAAGTGTGCTCTGAAGCTAG	1569
Db	1465	AATCCAGCAGTGTGCAGCCACTTGTGCACATTGTATGGGGCAAGTGTGCTCTGAAGCTAG	1524
Oy	1570	TGG	1572
Db	1525	TGG	1527

RESULT 6				
HUMMUCAB	HUMMUCAB	1904 bp	mRNA	linear PRI 07-JAN-1995
LOCUS	Human polymorphic epithelial mucin	(PEM)	mRNA, complete cds.	
DEFINITION	J0581			
ACCESSION	J0581.1	GI:188869		
VERSION	polymorphic epithelial mucin,			
KEYWORDS	polymorphic epithelial mucin,			
SOURCE	Homo sapiens adult adenocarcinoma CDNA to mRNA.			
ORGANISM	Homo sapiens			

REFERENCE	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1804)	Gendler, S.J., Lancaster, C.A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchett, J., Bemerson, L., Lalani, E.N. and Wilson, D.	Journal of Molecular Cloning and Expression of human tumor-associated polymorphic epithelial mucin	
2 (bases 1805 to 1890)	J. Biol. Chem. 265 (25), 15286-15293 (1990)		
3 (bases 1891 to 1980)			
4 (bases 1981 to 2070)			
5 (bases 2071 to 2160)			
6 (bases 2161 to 2250)			
7 (bases 2251 to 2340)			
8 (bases 2341 to 2430)			
9 (bases 2431 to 2520)			
10 (bases 2521 to 2610)			
11 (bases 2611 to 2700)			
12 (bases 2701 to 2790)			
13 (bases 2791 to 2880)			
14 (bases 2881 to 2970)			
15 (bases 2971 to 3060)			
16 (bases 3061 to 3150)			
17 (bases 3151 to 3240)			
18 (bases 3241 to 3330)			
19 (bases 3331 to 3420)			
20 (bases 3421 to 3510)			
21 (bases 3511 to 3600)			
22 (bases 3601 to 3690)			
23 (bases 3691 to 3780)			
24 (bases 3781 to 3870)			
25 (bases 3871 to 3960)			
26 (bases 3961 to 4050)			
27 (bases 4051 to 4140)			
28 (bases 4141 to 4230)			
29 (bases 4231 to 4320)			
30 (bases 4321 to 4410)			
31 (bases 4411 to 4500)			
32 (bases 4501 to 4590)			
33 (bases 4591 to 4680)			
34 (bases 4681 to 4770)			
35 (bases 4771 to 4860)			
36 (bases 4861 to 4950)			
37 (bases 4951 to 5040)			
38 (bases 5041 to 5130)			
39 (bases 5131 to 5220)			
40 (bases 5221 to 5310)			
41 (bases 5311 to 5400)			
42 (bases 5401 to 5490)			
43 (bases 5491 to 5580)			
44 (bases 5581 to 5670)			
45 (bases 5671 to 5760)			
46 (bases 5761 to 5850)			
47 (bases 5851 to 5940)			
48 (bases 5941 to 6030)			
49 (bases 6031 to 6120)			
50 (bases 6121 to 6210)			
51 (bases 6211 to 6300)			
52 (bases 6301 to 6390)			
53 (bases 6391 to 6480)			
54 (bases 6481 to 6570)			
55 (bases 6571 to 6660)			
56 (bases 6661 to 6750)			
57 (bases 6751 to 6840)			
58 (bases 6841 to 6930)			
59 (bases 6931 to 7020)			
60 (bases 7021 to 7110)			
61 (bases 7111 to 7200)			
62 (bases 7201 to 7290)			
63 (bases 7291 to 7380)			
64 (bases 7381 to 7470)			
65 (bases 7471 to 7560)			
66 (bases 7561 to 7650)			
67 (bases 7651 to 7740)			
68 (bases 7741 to 7830)			
69 (bases 7831 to 7920)			
70 (bases 7921 to 8010)			
71 (bases 8011 to 8100)			
72 (bases 8101 to 8190)			
73 (bases 8191 to 8280)			
74 (bases 8281 to 8370)			
75 (bases 8371 to 8460)			
76 (bases 8461 to 8550)			
77 (bases 8551 to 8640)			
78 (bases 8641 to 8730)			
79 (bases 8731 to 8820)			
80 (bases 8821 to 8910)			
81 (bases 8911 to 9000)			
82 (bases 9001 to 9090)			
83 (bases 9091 to 9180)			
84 (bases 9181 to 9270)			
85 (bases 9271 to 9360)			
86 (bases 9361 to 9450)			
87 (bases 9451 to 9540)			
88 (bases 9541 to 9630)			
89 (bases 9631 to 9720)			
90 (bases 9721 to 9810)			
91 (bases 9811 to 9900)			
92 (bases 9901 to 9990)			

by S.J.Gendler, 26-JUN-1990.
Location/Qualifiers
1. .1804

source	
1. .1804	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/map="Xq26.3-q27.1"
	/cell_line="BT20"
	/tissue_type="adenocarcinoma"
	/dev_stage="adult"
gene	1. .1804
	/gene="MCNAA"
mRNA	<1. .1804
	/gene="MCNAA"
	/product="PEM mRNA"
CDS	73. .1500

```

"/gene="MCNAA"
/codon_start=1
/product="polymorphic epithelial mucin"
/protein_id="AA59876.1"
/db_xref="GI:188870"
/tranlation="MPGQTSPPFLILLITVLTVTYVSGHASTGEGEKETAKORSS
VSGTSKXNAVMTSVLSHSPGSSPTQGGDTTLATPATERASSAKATQWDTVSP
VTRPAGSTTPPADYTSAPDNKRAPGSTAPAGVTSADVTRPAGSTAPAGVTS
ADNPAPAGSTAPVNAVTSAGSAGSAGSTAPVHNGTSAAVTTTPASKSTPFSIPSH
SDPTLLASHSTKTDASTHSTVTPFSGSHSTSPOLSTGVSPFLSTPHESTSN
SLEDSTVSYELORDISEMFLQIKQAGFLGSLNITKFRGVSVQVLTARLETNTV
HVEVTFQMYKTEAKRSYNNLTISDVSDVFPFSPAQSGAGPGWGLITLVGVVA
LAIIVTLAVACQCRKRTYGOLDIPPADYTHPMSEYPTVTHGRVYPPSTDSRPE
KTSANGCSLSYTNPAVATSANL"
73..135
sig_peptide

```

Feature	Value
sig_peptide	73..135 /gene="MCNNA"
mat_peptide	136..1497 /gene="MCNNA" /product="polymorphic epithelial mucin"
polya_signal	1783..1788 /gene="MCNNA"
BASE COUNT	361 a 622 c 434 g 387 t
ORIGIN	

Query Match	89.8%;	Score 1412.2;	DB 9;	Length 1804;
Best Local Similarity	95.3%;	Pred. No. 7e-248;		
Matches 1490; Conservative	0;	Mismatches 13;	Indels 60;	Gaps 1;

QY	10	GGTGTGCTGAATCTGTTCTGCGCCCTCCCAACCCATTTCAACACACATGACACCGGAC	69
Db	25	GGCTGGCTGAATCTGTTCTGCGCCCTCCCAACCCATTTCAACACACATGACACCGGAC	84
QY	70	ACCAAGTCACCTTTCTTCTGCGCTGCTCCCAAGCTTACAGTTTACAGTTCT	129
Db	85	ACCAAGTCACCTTTCTTCTGCGCTGCTCCCAAGCTTACAGTTTACAGTTCT	144
QY	130	GGTATGCAAGCTCTACCCCAAGGTGAGAAAGAGATTTGAGTACCCAGAAATTCA	189
Db	145	GGTATGCAAGCTCTACCCCAAGGTGAGAAAGAGATTTGAGTACCCAGAAATTCA	204
QY	190	GTGGCCAGCTCTACTGAGAGAAATGTGTGAGTATGACAGAGAGCTACTTCCAGCCAC	249
Db	205	GTGGCCAGCTCTACTGAGAGAAATGTGTGAGTATGACAGAGAGCTACTTCCAGCCAC	264
QY	250	AGCCCCGTTAAGGCTCTCTCAACACTCAGGAGAGAGATGTCACTCTGGCCCCGGCAG	309
Db	265	AGCCCCGTTAAGGCTCTCTCAACACTCAGGAGAGAGATGTCACTCTGGCCCCGGCAG	324
QY	310	GAACCAAGCTTCAAGTTCAAGTGTCACTCTGGGGAGAGAGATGTCACTCTGGTCCCAATCAC	369
Db	325	GAACCAAGCTTCAAGTTCAAGTGTCACTCTGGGGAGAGAGATGTCACTCTGGTCCCAATCAC	384
QY	370	AAGCCAGCTCTGGGCTTCAACCACTCCGCGACGCCAGATGTCACTCAAGCCCCGGACAA	429
Db	385	AAGCCAGCTCTGGGCTTCAACCACTCCGCGACGCCAGATGTCACTCAAGCCCCGGACAA	444
QY	430	AAGCCAGCTCTGGGCTTCAACCACTCCGCGACGCCAGATGTCACTCTGGCCCCGGACAA	489
Db	445	AAGCCAGCTCTGGGCTTCAACCACTCCGCGACGCCAGATGTCACTCTGGCCCCGGACAA	474
QY	490	AAGCCAGCTCTGGGCTTCAACCACTCCGCGACGCCAGATGTCACTCTGGCCCCGGACAA	549
Db	475	-----GCCACAGTGTCACTCTGGCCCCGGACAA	504
QY	550	AAGCCAGCTCTGGGCTTCAACCACTCCGCGACGCCAGATGTCACTCTGGCCCCGGACAA	609
Db	505	AAGCCAGCTCTGGGCTTCAACCACTCCGCGACGCCAGATGTCACTCTGGCCCCGGACAA	564
QY	610	AAGCCAGCTCTGGGCTTCAACCACTCCGCGACGCCAGATGTCACTCTGGCCCCGGACAA	669
Db	565	AAGCCAGCTCTGGGCTTCAACCACTCCGCGACGCCAGATGTCACTCTGGCCCCGGACAA	624
QY	670	GCATCAGGCTCACTCTGATGTGACAAAGGACCTCTGGCCAGGAGCTTACCAAAAC	729
Db	625	GCATCAGGCTCACTCTGATGTGACAAAGGACCTCTGGCCAGGAGCTTACCAAAAC	684
QY	730	CCACCCAGCAAGAGCACTCAACCCAGATTTCCAGCCACCACTCTGATATCTCTTACCAAC	789
Db	685	CCACCCAGCAAGAGCACTCAACCCAGATTTCCAGCCACCACTCTGATATCTCTTACCAAC	744
QY	790	CTTGCAAGCCATGACCAAGAGTATGACAGTATGACCTCAACATAGACAGGTACTCTCT	849
Db	745	CTTGCAAGCCATGACCAAGAGTATGACAGTATGACCTCAACATAGACAGGTACTCTCT	804
QY	850	CTCACCTCTCCCAATACAGCACTTCTCCCAAGTGTCTTACTGAGGCTCTCTTTCTTTTTC	909
Db	805	CTCACCTCTCCCAATACAGCACTTCTCCCAAGTGTCTTACTGAGGCTCTCTTTCTTTTTC	864
QY	910	CTGTCTTTTCAATTTCAAACTCTCAAGTTAATTTCTCTGGAAGATCCAGACAGAC	969
Db	865	CTGTCTTTTCAATTTCAAACTCTCAAGTTAATTTCTCTGGAAGATCCAGACAGAC	924
QY	970	TACATCAAGAGTGTGAGAGACATTTCTGAAATCTTTTTCAGATTTTAAACAAGGG	1029
Db	925	TACATCAAGAGTGTGAGAGACATTTCTGAAATCTTTTTCAGATTTTAAACAAGGG	984
QY	1030	GGTTTCTGGGCTCTCCAAATATTAAAGTTCAAGGCCAGATCTGTGTGTGTATCAATTGACT	1089

Db 985 GGTTCCTGGGCTCTCCAAATATTAAGTTGAGGCGAGAGATGTGGTGGTAAATGACT 1044
Oy 1090 CTGGCCCTCCGAGAAAGTACCAATCATGTCACAGAGCTGAGACACAGTTCAATCAGAT 1149
Db 1045 CTGGCTCTCCGAGAAAGTACCAATCATGTCACAGAGCTGAGACACAGTTCAATCAGAT 1104
Oy 1150 AAAACGAGAACAGCTCTGATATTAACCTGACGATCTCAGACGTCAGCGTGAATGATG 1209
Db 1105 AAAACGAGAACAGCTCTGATATTAACCTGACGATCTCAGACGTCAGCGTGAATGATG 1164
Oy 1210 CCATTTCCTTCTCTGCGCAAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1269
Db 1165 CCATTTCCTTCTCTGCGCAAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1224
Oy 1270 GTGCTGTCTGTGTCTGT 1329
Db 1225 GTGCTGTCTGTGTCTGT 1284
Oy 1330 CAGTGCCTGAGAAAGAACTACAGGAGCTGAGACATCTTTCAGCCGAGATACCTACAT 1389
Db 1285 CAGTGCCTGAGAAAGAACTACAGGAGCTGAGACATCTTTCAGCCGAGATACCTACAT 1344
Oy 1390 CCTATGAGCGAGTACCCCACTACCAACCAATGAGGCTGATGTGCTGCTTACAGATACC 1449
Db 1345 CCTATGAGCGAGTACCCCACTACCAACCAATGAGGCTGATGTGCTGCTTACAGATACC 1404
Oy 1450 GATGTGAGCCCTATGAGAGGTTCTGAGGATATGATGAGGAGAGAGCTCTTACACA 1509
Db 1405 GATGTGAGCCCTATGAGAGGTTCTGAGGATATGATGAGGAGAGAGCTCTTACACA 1464
Oy 1510 AAACGAGAGTGGAGGACCACTTCTGCAACTTTGTAAGGAGACGTCGCTCTGAGCTGAG 1569
Db 1465 AAACGAGAGTGGAGGACCACTTCTGCAACTTTGTAAGGAGAGAGTCGCGGAGAGCTGAG 1524
Oy 1570 TGG 1572
Db 1525 TGG 1527

RESULT 7
GIBMUC1A 1795 bp DNA linear PRI 17-JUN-1995
LOCUS Hylobates lar epithelial mucin (MUC1) gene, complete cds.
DEFINITION
L41589
ACCESSION
L41589.1 GI:897611
VERSION
MUC1 gene; epithelial mucin; membrane protein; mucin; variable
KEYWORDS
number tandem repeat.
SOURCE
Hylobates lar (clone library: cos203 cosmid library) adult DNA.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
REFERENCE
Spicer, A.P., Dubig, T., Chilton, B.S. and Gendler, S.J.
AUTHORS
1 (bases 1 to 1795)
TITLE
Analysis of mammalian MUC1 genes reveals potential functionally
JOURNAL
Genomics (1995) In press
COMMENT
On Jul 14, 1995 this sequence version replaced gi:886240.
The gibbon sequence presented here contains two consensus 60 bp
tandem repeat units. From studies of the human MUC1 gene, a single
allele can contain from 20 up to 125 repeat units. The gibbon MUC1
clones we describe here contained approximately 35 tandem repeats.
FEATURES
Location/Qualifiers
1..1795
/organism="Hylobates lar"
/db_xref="taxon:9580"
/clone_lib="cos203 cosmid library"
/dev_stage="adult"
1..1795
/gene="MUC1"
1..1795
/gene="MUC1"
/product="epithelial mucin"

CDS
69..1496
/gene="MUC1"
/note="1287..1493 cytoplasmic domain, 1194..1286
transmembrane domain"
/codon_start=1
/product="epithelial mucin"
/protein_id="AA6965.1"
/db_xref="GI:897612"
/translation="MTPTGQSLFLLILLTVLTVTYVSGHASTPGEKETASORSS
MBSSTKRYKVENTSSVLSHSPSGSSSTGQDVSLAPETPAGSAATWGDVTP
VIRPAPGSTSPAGVTSAPDTRPALGSTAPYKHVTSAPDTRPALGSTAPYKHVTS
ADPTIRPLGSTAPYKHVTSAPDTRPALGSTAPYKHVTSAPDTRPALGSTAPYKHVTS
SDPTPLTSHSTKTDASTHSTVSPSTSNSTSPQSLVSPFUSFPIISLQNS
SLDPTNLYOELORDISELILQYKQDPLVSNIKRPSVVOGSLTLPFRGTNN
HDVEAFQNHKEAASRYNLISDVSVDVPPFPAQSGAGVPMGIALLVLCVLA
LATVYLALAVCCRRKRYGQDLPAPDAVHPMSEYPTVHTHRVYVPSTWRSFYE
KYSBNGSGSLSYNPAPVATSNLV"
69..131
/gene="MUC1"
453..572
/rpt_type=tandem
/rpt_unit=453..512
/rpt_unit=513..572
453..512
/gene="MUC1"
513..572
/gene="MUC1"
1774..1779
/gene="MUC1"
BASF COUNT 365 a 429 g 390 t
ORIGIN
Query Match 84.5%; Score 1328.4; DB 9; Length 1795;
Best Local Similarity 91.9%; Pred. No. 1.4e-232;
Matches 1438; Conservative 0; Mismatches 66; Indels 60; Gaps 1;
Oy 9 TGACTGCTTGAATCTGTTCTGCCCCCTCCCAACCACTTTCACCAACCAACCAACCGG 68
Db 20 TGCTTCTGAGTCTGTTCTGCCCCCTCCCAACCACTTTCACCAACCAACCAACCGG 79
Oy 69 CACCCAGTCTCTTCTTCTTCTGCGGAGTCTCTCTCAAGTCTCAAGTCTCAAGTCTC 128
Db 80 CACCCAGTCTCTTCTTCTTCTGCGGAGTCTCTCTCAAGTCTCAAGTCTCAAGTCTC 139
Oy 129 TGCTATGCAAGCTTCAACCCAGGAGGAGAAAGAGACCTTGCTTACCAAGAAATTTC 188
Db 140 CGGTATGCAAGCTTCAACCCAGGAGGAGAAAGAGACCTTGCTTACCAAGAAATTTC 199
Oy 189 AGTGCAGAGCTCTACTGAGAAAGATGCTGAGATGACACAGAGCTACTCTCCAGCA 248
Db 200 AATGCCAGCTTCTACTGAGAAAGATGCTGAGATGACACAGAGCTACTCTCCAGCA 259
Oy 249 CAGCCCCGTTGAGGCTCTTCCACACTCAAGGAGACAGATGTCACTTGCCCCGCGAC 308
Db 260 CAGCCCCGTTGAGGCTCTTCCACACTCAAGGAGACAGATGTCACTTGCCCCGCGAC 319
Oy 309 GGAACAGCTTCAAGTCTCAAGTCTCACTTGGGAGACAGATGTCACTTGCCCCGCGAC 368
Db 320 GGAACAGCTTCAAGTCTCAAGTCTCACTTGGGAGACAGATGTCACTTGCCCCGCGAC 379
Oy 369 CAGGCCAGCTTGGGCTCTTCCACACCCGCGACAGCCAGATGTCACTTGCCCCGCGAC 428
Db 380 CAGGCCAGCTTGGGCTCTTCCACACCCGCGACAGCCAGATGTCACTTGCCCCGCGAC 428
Oy 429 CAGGCCAGCTTGGGCTCTTCCACACCCGCGACAGCCAGATGTCACTTGCCCCGCGAC 488
Db 429 CAGGCCAGCTTGGGCTCTTCCACACCCGCGACAGCCAGATGTCACTTGCCCCGCGAC 439
Oy 489 CAGGCCAGCTTGGGCTCTTCCACACCCGCGACAGCCAGATGTCACTTGCCCCGCGAC 548
Db 440 CAGGCCAGCTTGGGCTCTTCCACACCCGCGACAGCCAGATGTCACTTGCCCCGCGAC 499
Oy 549 CAGGCCAGCTTGGGCTCTTCCACACCCGCGACAGCCAGATGTCACTTGCCCCGCGAC 608

D	b	489	-----	488
Q	y	601	CCGGAACAAGGCGGCTTTGGGCTTCAACCGGCCCTCAAGTCCAAATATGCAACTTCGAGC	660
D	b	489	-----NAGGCGGNNNTTGGGCTTCAACCGGCCCTCAAGTCCAAATATGCAACTTCGAGC	540
Q	y	661	TCAGGCTCTGAGTCAAGGCTCAGCTTTCTACTCTGGGAGCAACGGGCGCTCTGCGAGGGCT	720
D	b	541	TCAGGCTCTGAGTCAAGGCTCAGCTTTCTACTCTGGGAGCAACGGGCGCTCTCTGCGAGGGCT	600
Q	y	721	ACCAACAACCCAGCGACGAAGAGCACTCAACCGATTTCCAGGCCACACTCTGATACT	780
D	b	601	ACCAACAACCCAGCGACGAAGAGCACTCAACCGAATTTCCAGGCCACAACCTCTGATACT	660
Q	y	781	CCTAACACCTTGGCAGCCATAGACAACAAGCTAGTATGCTGACCTCAACCATAGACG	840
D	b	661	CCTAACACCTTGGCAGCCATAGACAACAAGCTAGTATGCTGACCTCAACCATAGACG	720
Q	y	841	GTACCTCTCTCAACCTCTCAATCAAGCACTTCCCAAGTTGCTACTGAGGCTCTCT	900
D	b	721	GTACCTCTCTCAACCTCTCAATCAAGCACTTCCCAAGTTGCTACTGAGGCTCTCT	780
Q	y	901	TTCTTTTTCCTGTCTTTTCACATTTCAAACTCCAGTTTAATTCCTCTCTGGAAGATCC	960
D	b	781	TTCTTTTTCCTGTCTTTTCACATTTCAAACTCCAGTTTAATTCCTCTCTGGAAGATCC	840
Q	y	961	AGACCGCACTCTACCAAGAGCTGCGAGAGACAATTTCTGAATGTGTTTGGAGATTAT	1020
D	b	841	AGACCGCACTCTACCAAGAGCTGCGAGAGACAATTTCTGAATGTGTTTGGAGATTAT	900
Q	y	1021	AAACAAGGGGGTTTTCTGGGCTCTCCAAATTTAAGTTCAAGGCGAGATCTGTGGTGGTA	1080
D	b	901	AAACAAGGGGGTTTTCTGGGCTCTCCAAATTTAAGTTCAAGGCGAGATCTGTGGTGGTA	960
Q	y	1081	CAATTGACTCTGGGCTTCCAGAGAAGTACATCAATGTCCACACGTGAGACAAGTTTC	1140
D	b	961	CAATTGACTCTGGGCTTCCAGAGAAGTACATCAATGTCCACACGTGAGACAAGTTTC	1020
Q	y	1141	AATCAGTATATAAAGGAAGAGAGGCTCTGATATATACTGACGATCTCAAGCTCAAGGTG	1200
D	b	1021	AATCAGTATATAAAGGAAGAGAGGCTCTGATATATACTGACGATCTCAAGCTCAAGGTG	1080
Q	y	1201	AGTCATGTGCAATTTCTTTCTCTGCGCCAGTCTGAGGCTGAGGCTGCGAGGCGATC	1260
D	b	1081	AGTCATGTGCAATTTCTTTCTCTGCGCCAGTCTGAGGCTGAGGCTGCGAGGCGATC	1140
Q	y	1261	GGGCTGCTGGGCTCTGCTGATGTTCTGGTGTGGGCGCATGTGTATCTCATTTGSCCTG	1320
D	b	1141	GGGCTGCTGGGCTCTGCTGATGTTCTGGTGTGGGCGCATGTGTATCTCATTTGSCCTG	1200
Q	y	1321	GGGTCTGTCAAGTGCAGCGGCAAAAGACTACGAGGCACTCTTTCAAGCCCGGAGAT	1380
D	b	1201	GGGTCTGTCAAGTGCAGCGGCAAAAGACTACGAGGCACTCTTTCAAGCCCGGAGAT	1260
Q	y	1381	ACCTTACCAATTCATAGAGAGTATCCCACTACACAACCCATGAGGCGCTATATGCCCTT	1440
D	b	1261	ACCTTACCAATTCATAGAGAGTATCCCACTACACAACCCATGAGGCGCTATATGCCCTT	1320
Q	y	1441	AGCAGTACCGATCGTACGCCCTTATGAGAGGTTTTCTGAGAGTAAATGTGTGCGACGCTCT	1500
D	b	1321	AGCAGTACCGATCGTACGCCCTTATGAGAGGTTTTCTGAGAGTAAATGTGTGCGACGCTCT	1380
Q	y	1501	TCCTTACCAAAACCAAGAGTGTGGAGCACTTCTGCAACTGTATGAGGGGACGCTCGCCTC	1560
D	b	1381	TCCTTACCAAAACCAAGAGTGTGGAGCACTTCTGCAACTGTATGAGGGGACGCTCGCCTC	1440
Q	y	1561	TGAGCTGAAGTG	1572
D	b	1441	TGAGCTGAAGTG	1452

AF423030	1834 bp	mRNA	linear	SYN 11-MAR-2002
LOCUS	AF423030	Synthetic construct Homo sapiens mucin variant MUC-1CT80 (MUC1)		
DEFINITION		mRNA, complete cds; alternatively spliced.		
ACCESSION	AF423030			
VERSION	AF423030.1	GI:19338619		
KEYWORDS				
SOURCE		synthetic construct.		
ORGANISM		synthetic construct.		
REFERENCE		1 (bases 1 to 1834)		
AUTHORS		Hinojosa-Kurtzberg, A.M., Johansson, M.J., Madsen, C.S., Hansson, G.C.		
TITLE		Mucin MUC1 is part of the mucus accumulated in cystic fibrosis mice		
JOURNAL		2 (baaes 1 to 1834)		
REFERENCE		Hinojosa-Kurtzberg, A.M. and Gendler, S.J.		
AUTHORS		Submitted (21-SEP-2001) Biochemistry, Mayo Clinic, 13400 E. Shea		
JOURNAL		Boulevard, Scottsdale, AZ 85259, USA		
FEATURES		Location/Qualifiers		
source		1. 1834		
		/organism="synthetic construct"		
		/db_xref="taxon:32630"		
		/note="Homo sapiens gene in transgenic Mus musculus		
		C57BL/6; isolated from intestinal mucosa"		
gene		1. 1834		
		/gene="MUC1"		
		1. 1452		
		/gene="MUC1"		
		/note="alternatively spliced; contains exon 6a resulting		
		in variant hydrophilic tail of carboxy-terminal domain;		
		lacks sites for beta-catenin and Grb2 interactions;		
		derived from Homo sapiens"		
		/codon_start=1		
		/transl_table=11		
		/product="mucin variant MUC-1CT80"		
		/protein_id="AAU6734.1"		
		/db_xref="GI:19338620"		
		/translation="MTRPGTSPFFLLLLVYLVTGSGHASTPGGEKETSATQRSS		
		VPSESTKNAVSWTSSVLSSHPGSGSTYQGDQVTLPAPEHPSGSAATWQDVTVP		
		AGTGGATTTPADHTVAPDNKPAFGASTAPAGHTGATTPAGSTAPAGHTVS		
		APNRPALGSTAPPVHNVTASGASGASTLHNHNSAATTPASSTPSLPSHH		
		SPRTPLSTKTDASTSTHSTVPSPTLSNHSSTPOLSTGVSFFLSPHSLNQS		
		SLDEPSTDYQELORDISSEMFIDTIYKGGSTPLSNIKRFGSVYFQTLARECTIV		
		HDVEFPNDYKTEASRNLTISDVSSVPEPPSAGSGAGVGMKALDLYNCVLA		
		LAIVLYIAQCCRRKNYGLDIPPARDTNAPSPSEPTVYHGRVPPSSIDSPEY		
		KVRLGPTGGKCRRWLKKDSBGGTMTQDAWKR"		
BASE COUNT	394 a	570 c	488 g	382 t
ORIGIN				
Query Match	80.7%	Score 1269;	DB 12;	Length 1834;
Best Local Similarity	95.1%;	Pred. No. 9.4e-222;		
Matches 1345;	Conservative	0;	Mismatches 10;	Indels 60;
			Gaps 1;	
58	ATGACACCGGGGACCCAGTCTCTCTTCTCTCTGCTGCTCTCTCAAGTCTTACAGTT	117		
Db	1 ATGACACCGGGGACCCAGTCTCTCTTCTCTCTGCTGCTCTCTCAAGTCTTACAGTT	60		
118	GTTACAGGTTCTGTCATGCAACCTCTAACCCAGAGTGAAGAAAGAAAGACTTGGGGTACC	177		
Db	61 GTTACAGGTTCTGTCATGCAACCTCTAACCCAGAGTGAAGAAAGAAAGACTTGGGGTACC	120		
178	CAGAGAAAGTTCAAGTCCCAAGCTCTTACTAGAGAAAGATCTGTAGATGACACGAGCGTA	237		
Db	121 CAGAGAAAGTTCAAGTCCCAAGCTCTTACTAGAGAAAGATCTGTAGATGACACGAGCGTA	180		
238	CTCTCCAGCCACAGCCCGGTTAGGCTCTCCACCACTCTCAGGAGACAGAAATGTCACCTTG	297		
Db	181 CTCTCCAGCCACAGCCCGGTTAGGCTCTCTCCACCACTCTCAGGAGACAGAAATGTCACCTTG	240		
298	GCCCGGCGACGAGAAACAGCTTCAAGTTCAAGTCCCACTGCGGAGACAGAAATGTCACCTTG	357		

D	b	241	GGCCCCGGACAAGSAAACAAGTTTCAAGTTGACGTGCAACCTGAGGAGACAAGATGTCACTCTCG	300
O	y	358	GTCCCAAGTCAACAGGCGCAGGCTCGGGCTCAACAACCGGCAAGCCCAAGATGTCACTCA	417
D	b	301	GTCCACAGTACCAAGGCAAGGCTCGGGCTCAACAACCGGCAAGCCCAAGATGTCACTCA	360
O	y	418	GGCCCGGACAAACAAGCCAGGCTCCGGGCTCAACCGGCCCCCGAGGATGTCACTCG	477
D	b	361	GGCCCGGACAAACAAGCCAGGCTCCGGGCTCAACCGGCCCCCA-----	402
O	y	478	GGCCCGGACACAGGCGCGCCCGGGCTCAACGGCCCCGGCGCCAGGTCACTCG	537
D	b	403	-----GGCAGCGGTCACTCG	420
O	y	538	GGCCCGGACACAGGCGCGCCCGGGCTCAACGGCCCCGGCTCAAGGTCACTCG	597
D	b	421	GGCCCGGACACAGGCGCGCCCGGGCTCAACGGCCCCGAGCCATGATGTCACTCG	480
O	y	598	GGCCCGGACAAAGGCGGAGCTTGAGGCTCAACGGGCTCTCAAGTCACAATGTCACTCG	657
D	b	481	GGCCCGGACAAAGGCGCGGCTTGAGGCTCAACGGGCTCTCAAGTCACAATGTCACTCG	540
O	y	658	GGCTGAGCTGTGATATAGGCTGAGCTTACTCTGTGAGACAACGGGCACTCTGCGAG	717
D	b	541	GGCTGAGCTGTGATATAGGCTGAGCTTACTCTGTGAGACAACGGGCACTCTGCGAG	600
O	y	718	GCTACACAACCCCAAGCCAGAGAGAGACTCAACCGAGATTCGACGCCACACTGTAT	777
D	b	601	GCTACACAACCCCAAGCCAGAGAGAGACTCAATTCGAGCCACACTGTAT	660
O	y	778	ACTCTACCAACCTTGCGAGGCAATGACACAAGATGTGCAAGTCACTACATATAGC	837
D	b	661	ACTCTACCAACCTTGCGAGGCAATGACACAAGATGTGCAAGTCACTACATATAGC	720
O	y	838	AGGTAACCTTCCTCACTCTCTCAATCAAGACTTTCGCCAGTTGTCTAGAGGATC	897
D	b	721	AGGTAACCTTCCTCACTCTCTCAATCAAGACTTTCGCCAGTTGTCTAGAGGATC	780
O	y	898	TCTTTCTTTTCTGTCTTTTCAATTTCAAACTCGAGTTAATTCCTCTGAGAT	957
D	b	781	TCTTTCTTTTCTGTCTTTTCAATTTCAAACTCGAGTTAATTCCTCTGAGAT	840
O	y	958	CCGAGCAAGCTACCTACCAAGAGCTGCAAGAGACATTTGAAATGTTTTGACATTT	1017
D	b	841	CCGAGCAAGCTACCTACCAAGAGCTGCAAGAGACATTTGAAATGTTTTGACATTT	900
O	y	1018	TATTAACAAGGGGGTTTTCTGAGGCTCTCAATATTAAGTCAAGGCAAGATCTGTGTG	1077
D	b	901	TATTAACAAGGGGGTTTTCTGAGGCTCTCAATATTAAGTCAAGGCAAGATCTGTGTG	960
O	y	1078	GTAACAATTGACTGTGAGCTTCCGAGAGAGTACCAATGTCCAGACGTGAGACAG	1137
D	b	961	GTAACAATTGACTGTGAGCTTCCGAGAGAGTACCAATGTCCAGACGTGAGACAG	1020
O	y	1138	TTCAATCAAGTATTAACAAGAGGAGCTCTCAATTTAACTGACGATCTAGAGCTGAGC	1197
D	b	1021	TTCAATCAAGTATTAACAAGAGGAGCTCTCAATTTAACTGACGATCTAGAGCTGAGC	1080
O	y	1198	GTGAGTCAATGTGCAATTTCTTTTCTGTGCCAGTCTGGGGCTGGGATGTCCAGCTGGGC	1257
D	b	1081	GTGAGTCAATGTGCAATTTCTTTTCTGTGCCAGTCTGGGGCTGGGATGTCCAGCTGGGC	1140
O	y	1258	ATGCGCTGCTGATCTGTGTCTGTGTTCTGTGTTGCGCTGAGCAATGTCTATCTCATTTGCC	1317
D	b	1141	ATGCGCTGCTGATCTGTGTCTGTGTTCTGTGTTGCGCTGAGCAATGTCTATCTCATTTGCC	1200
O	y	1318	TTGGAGTGTGTGACATGTGCGCGCAAAACCTACGGGCAAGCTGACATCTTTCAGCGCGG	1377
D	b	1201	TTGGAGTGTGTGACATGTGCGCGCAAAACCTACGGGCAAGCTGACATCTTTCAGCGCGG	1260
O	y	1378	GATATCAACATCTATAGAGGAGTACCCACCTTACACATCCATGGGGCTATGTGTCC	1437
D	b	1261	GATATCAACATCTATAGAGGAGTACCCACCTTACACATCCATGGGGCTATGTGTCC	1320

[illegible]

Oy	238	CTCTCCAGCAGCAGACCCCGGTTTCAGGCTCCTCCACACATAGGAGACAGGATGTCACTCTG	297
Db	181	CTCTCCAGCAGCAGACCCCGGTTTCAGGCTCCTCCACACATAGGAGACAGGATGTCACTCTG	240
Oy	298	GCCCCGGCCACGGAAACAGCTTCAGGTTCACTGTCACCTGGGAGACAGATGTCACTCTG	357
Db	241	GCCCCGGCCACGGAAACAGCTTCAGGTTCACTGTCACCTGGGAGACAGATGTCACTCTG	300
Oy	358	GTTCCAGTCAACAGGACAGGCTCTGGGCTTCACACACCCCGCAGCCACAGATGTCACTCA	417
Db	301	GTTCCAGTCAACAGGACAGGCTCTGGGCTTCACACACCCCGCAGCCACAGATGTCACTCA	360
Oy	418	GCCCCGGAGAAACAGGACCCCGGGCTCAACCGCCCGCCGGCCAGGATGTCACTCTG	477
Db	361	GCCCCGGAGAAACAGGACCCCGGGCTCAACCGCCCGCCGGCCAGGATGTCACTCTG	402
Oy	478	GCCCCGGAGACCAAGGACCCCGGGCTCAACCGCCCGCCGGCCAGGATGTCACTCTG	537
Db	403	-----GCCACGGATCACTCTG-----GCCACGGATCACTCTG	420
Oy	538	GCCCCGGAGACCAAGGACCCCGGGCTCAACCGCCCGCCGGCCAGGATGTCACTCTG	597
Db	421	GCCCCGGAGACCAAGGACCCCGGGCTCAACCGCCCGCCGGCCAGGATGTCACTCTG	480
Oy	598	GCCCCGGAGACAGGCTGGGCTTGGGCTCAACCGCCCGCCAGTGTCAAAATGTCACTCTG	657
Db	481	GCCCCGGAGACAGGCTGGGCTTGGGCTCAACCGCCCGCCAGTGTCAAAATGTCACTCTG	540
Oy	658	GCCCTCAGGCTGTGCATCAGGCTCAGCTTCACTGTGTGCAACACGGCACTCTGCGAG	717
Db	541	GCCCTCAGGCTGTGCATCAGGCTCAGCTTCACTGTGTGCAACACGGCACTCTGCGAG	600
Oy	718	GCTACCAACAACCCCGAGGACAGAGACATCCACCCAGATTTCCAGGCCACACTGTAT	777
Db	601	GCTACCAACAACCCCGAGGACAGAGACATCCATTTCTCAATTCGAGGCCACACTGTAT	660
Oy	778	ACTCTCTACACCCCTTGCAGCCATAGCACCAAGACTGATGTCAGATGACATCAACATGAC	837
Db	661	ACTCTCTACACCCCTTGCAGCCATAGCACCAAGACTGATGTCAGATGACATCAACATGAC	720
Oy	838	ACGGTACCTCTCTCAGCTCTCTCCCAATCAGAGCACTTCTCCGAGTTGTCTAGGGATC	897
Db	721	ACGGTACCTCTCTCAGCTCTCTCCCAATCAGAGCACTTCTCCGAGTTGTCTAGGGATC	780
Oy	898	TCTTTCTTTTCCGTCTTTTCAATTTCAAACCTCCAGTTAATCTCTCGAGAAAT	957
Db	781	TCTTTCTTTTCCGTCTTTTCAATTTCAAACCTCGATTATTTCTCTCGAGAAAT	840
Oy	958	CCGAGCACCGACTACTACCAAGCTCAGAGAGACATTTCTGAAATGTTTTGCAATTT	1017
Db	841	CCGAGCACCGACTACTACCAAGCTCAGAGAGACATTTCTGAAATGTTTTGCAATTT	900
Oy	1018	TATTAACAAGGGGGTTTTCTGGGCTCTCCAAATTAAGTTCAAGCCAGGATCTGTGTG	1077
Db	901	TATTAACAAGGGGGTTTTCTGGGCTCTCCAAATTAAGTTCAAGCCAGGATCTGTGTG	960
Oy	1078	GTAACAATTGACTTGGGCTTCGAGAGAGATACCAATATGTCAACGACGTGGAGACAG	1137
Db	961	GTAACAATTGACTTGGGCTTCGAGAGAGATACCAATATGTCAACGACGTGGAGACAG	1020
Oy	1138	TTTCATCAGTATTAACCGAGACAGCTCTTGATATTAACCTGACGATCTCAAGCTCAGC	1197
Db	1021	TTTCATCAGTATTAACCGAGACAGCTCTTGATATTAACCTGACGATCTCAAGCTCAGC	1080
Oy	1199	GTGAGTATGTGCAATTTCTTTCTTCGACAGCTCGGGGCTGGGGGTCAGAGCTGAGGC	1257
Db	1081	GTGAGTATGTGCAATTTCTTTCTTCGACAGCTCGGGGCTGGGGGTCAGAGCTGAGGC	1140
Oy	1258	ATGCGCTGCTGGTGTGCTGTGTGTCTTGGTTGGCTGGCCATTTGTCTATCTCAATTTGC	1317
Db	1141	ATGCGCTGCTGGTGTGCTGTGTGTCTTGGTTGGCTGGCCATTTGTCTATCTCAATTTGC	1200

	QY	1318	TTCGTGCTGTCAAGTGCACCGGAAGAATCAGGGAGACTGGAATCTTTCCAGCCGG	1377
Db	1201	TTGGCTGCTGTCAAGTGCACCGGAAGAATCAGGGAGACTGGAATCTTTCCAGCCGG	1260	
QY	1378	GATACCTTACCATCTTAGTGAGCGATGCCA CCTACCAACCATTGGGCGCTATATGCCCC	1437	
Db	1261	GATACCTTACCATCTTAGTGAGCGAGTACCCCACTTACCAACCATTGGGCGCTATATGCCCC	1320	
QY	1438	CCTAGCAGTACCGATCGTAGCCCTTAGAGAAGG	1471	
Db	1321	CCTAGCAGTACCGATCGTAGCCCTTAGAGAAGG	1354	
RESULT 11				
AR030811				
LOCUS	AR030811	6192 bp	DNA	linear
DEFINITION	Sequence 1 from patent US 5861381.			PAT 29-SEP-1999
ACCESSION	AR030811			
VERSION	AR030811.1	GI:5944025		
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 6192)			
TITLE	Chambon, P., Kleny, M.-P., Lathe, R. and Harevenvi, M.			
JOURNAL	Pharmaceutical composition for the treatment of a malignant tumor			
FEATURES	Patent: US 5861381-A 1 19-'04N-1999;			
source	Location/Qualifiers			
	1..6192			
	/organism="unknown"			
BASE COUNT	695 a 2628 c 1513 g 636 t		720 others	
ORIGIN				
Query Match	76.1%;	Score 1196.6;	DB 6;	Length 6192;
Best Local Similarity	92.2%;	Pred. No. 1.5e-208;		
Matches 1233; Conservative	0;	Mismatches 104;	Indels 0;	Gaps 0
QY	246	CCACAGCCCCGGATTACAGGCTCTCCACACA CTTCAGGGA CAGATGTCACTTGGCCCGGCG	305	
Db	4866	CNNNAGCGCNNNCGGGCTTCACCGCCCCCNMNGCCACGGTGTCACTTCGGCCCCGGA	4925	
QY	306	CACGGAACGAGTTCAGGTTCACTGTCGCACTGGGGAA CAGGATGTCACTTCGGTCCAGT	365	
Db	4926	CNNNAGCGCNNNCGGGCTTCACCGCCCCCNMNGCCACGGTGTCACTTCGGCCCCGGA	4985	
QY	366	CACGAGCGAGCCCTGGGCTTCACACCCCAGACCCA GATGTCACTTCAGCCCCGGA	425	
Db	4986	CNNNAGCGCNNNCGGGCTTCACCGCCCCCNMNGCCACGGTGTCACTTCGGCCCCGGA	5045	
QY	426	CAA CAAGCACGAGCCCCGGGCTTCACCGCCCCCGGAGCCCA CGGTGTCACTTCGGCCCCGGA	485	
Db	5046	CNNNAGCGCNNNCGGGCTTCACCGCCCCCNMNGCCCA CGGTGTCACTTCGGCCCCGGA	5105	
QY	486	CACGAGCGCCCCCGGGCTTCACCGCCCCCGGCGCCAGAGGTATCACTTCGGCCCCGGA	545	
Db	5106	CNNNAGCGCNNNCGGGCTTCACCGCCCCCNMNGCCACGGTGTCACTTCGGCCCCGGA	5165	
QY	546	CACCAAGCGCGGCCCGGGCTTCACCGCCCCCGGAGCCCA GGTGTCACTTCGGCCCCGGA	605	
Db	5166	CNNNAGCGCNNNCGGGCTTCACCGCCCCCNMNGCCCA CGGTGTCACTTCGGCCCCGGA	5225	
QY	606	CAACAGGCGGGGCTTGGGCTTCACCGCCCCCTCGAGTCCACAATGTCACTTCGGCTTCAGG	665	
Db	5226	CNNNAGCGANNNTTGGGCTTCACCGCCCCCTCGAGTCCACAATGTCACTTCGGCTTCAGG	5285	
QY	666	CTCTGCATCAGGCTCAAGCTTTACTCTGTGTGCA CAA CGGCA CTTGTGCCAGGGCTTACAC	725	
Db	5286	CTCTGCATCAGGCTCAAGCTTTACTCTGTGTGCA CAA CGGCA CTTGTGCCAGGGCTTACAC	5345	
QY	726	AAACCCAGCGAGAGAGACTTCACACCCAGACATTCAC CCACCACTCTGATACTTCTTAC	785	
Db	5346	AAACCCAGCGAGAGAGACTTCACACCCAGACATTCAC CCACCACTCTGATACTTCTTAC	5405	

QY 786 CACCTTGGCAGCAGTACGACCAAGCTGATGCGAGTACGCTACCATACAGGTTAC 845
 Db 5406 CACCTTGGCAGCAGTACGACCAAGCTGATGCGAGTACGCTACCATACAGGTTAC 5465
 QY 846 TCCTCTGACCTCTCCATACAGCACTTCTCCAGTTGTCTACTGGGGTCTCTTCTT 905
 Db 5466 TCCTCTGACCTCTCCATACAGCACTTCTCCAGTTGTCTACTGGGGTCTCTTCTT 5525
 QY 906 TTTCTGTCTTTTCAATTCAATTAACCTCAGATTAAATTCCTCTGAGAAATCCAGAC 965
 Db 5526 TTTCTGTCTTTTCAATTCAATTAACCTCAGATTAAATTCCTCTGAGAAATCCAGAC 5585
 QY 966 CGACTTACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTGCAATTTATAACA 1025
 Db 5586 CGACTTACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTGCAATTTATAACA 5645
 QY 1026 AGGGGGTTTTCTGGGCTCTCCATATTAATGTTAGGCGCAGATCTGAGTACAT 1085
 Db 5646 AGGGGGTTTTCTGGGCTCTCCATATTAATGTTAGGCGCAGATCTGAGTACAT 5705
 QY 1086 GACTCTGGCTTCCGAGAGTACATCATATGTCACAGCTGAGACACAGTTCAATCA 1145
 Db 5706 GACTCTGGCTTCCGAGAGTACATCATATGTCACAGCTGAGACACAGTTCAATCA 5765
 QY 1146 GTATTAACCGAAGCAGCTCTCGATATTAACCTGACATCTCAGAGTCAAGTGAATCA 1205
 Db 5766 GTATTAACCGAAGCAGCTCTCGATATTAACCTGACATCTCAGAGTCAAGTGAATCA 5825
 QY 1206 TGTGCAATTTCTCTTCTGCGCAGCTGAGGCTGGGGTCCAGGCTGGGGATTCGGCT 1265
 Db 5826 TGTGCAATTTCTCTTCTGCGCAGCTGAGGCTGGGGTCCAGGCTGGGGATTCGGCT 5885
 QY 1266 GCTGTGTGCTGTGTGTCTGTGTGCGCTGGCATTGTCTATCTCATTTGCGCTGT 1325
 Db 5886 GCTGTGTGCTGTGTGTCTGTGTGCGCTGGCATTGTCTATCTCATTTGCGCTGT 5945
 QY 1326 CTGTGAGTGGCGCGGAAAGACTAGAGGAGCTGGAATTTCCAGCCCGAGATACCTA 1385
 Db 5946 CTGTGAGTGGCGCGGAAAGACTAGAGGAGCTGGAATTTCCAGCCCGAGATACCTA 6005
 QY 1386 CCATCTTATGAGCAGTACCCCACTTACACACCCATGAGCGCTATGTGCCCCCTAGAC 1445
 Db 6006 CCATCTTATGAGCAGTACCCCACTTACACACCCATGAGCGCTATGTGCCCCCTAGAC 6065
 QY 1446 TACCGATGTAGCCCTATGAGAGGTTTCTGACAGTAATGTGTGAGAGCCCTCTTA 1505
 Db 6066 TACCGATGTAGCCCTATGAGAGGTTTCTGACAGTAATGTGTGAGAGCCCTCTTA 6125
 QY 1506 CACAAACCCAGAGTGGCAGCCACTTCTGCCAATTTGTAGGGGCACTGCCCCCTGAGC 1565
 Db 6126 CACAAACCCAGAGTGGCAGCCACTTCTGCCAATTTGTAGGGGCACTGCCCCCTGAGC 6185
 QY 1566 TGAGTGG 1572
 Db 6186 TGAGTGG 6192

RESULT 12
 ARI42537
 LOCUS ARI42537 6192 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 1 from patent US 6203795.
 ACCESSION ARI42537
 VERSION ARI42537.1 GI:15103823
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 6192)
 AUTHORS Chambon, B.; Kieny, M. P.; Iathe, R. and Hareuveni, M.
 TITLE Pharmaceutical Composition for the treatment or prevention of a
 malignant tumor
 JOURNAL Patent: US 6203795-A 1 20-MAR-2001;
 FEATURES Location/Qualifiers

source 1..6192
 BASE COUNT 695 a 2628 c 1513 g 636 t 720 others
 ORIGIN
 Query Match 76.1%; Score 1196.6; DB 6; Length 6192;
 Best Local Similarity 92.2%; Pred. No. 1.5e-208;
 Matches 1223; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
 QY 246 CCACAGGCCCCGTTACAGCTCTCCACACTCAGGAGCAGATGTCACTGCGCCCGG 305
 Db 4866 CANNAGGCGANNCCGGGCTCCACCGCCCGANNNGCCACGGTGTCACTCGGCGCGA 4925
 QY 306 CACGAAACGAGCTTCAAGTTCAAGTGCACCTGAGGAGACAGATGTCACTCGTCCAGT 365
 Db 4926 CANNAGGCGANNCCGGGCTCCACCGCCCGANNNGCCACGGTGTCACTCGGCGCGA 4985
 QY 366 CACAGGCGAGCCCTGGGCTCCACACCGCCCGCAGCCAGATGTCACTGAGCCCGGA 425
 Db 4986 CANNAGGCGANNCCGGGCTCCACCGCCCGANNNGCCACGGTGTCACTGAGCCCGGA 5045
 QY 426 CAACAGCCAGCCCGGGCTCCACCGCCCGCCCGCCAGGTGTCACTGAGCCCGGA 485
 Db 5046 CANNAGGCGANNCCGGGCTCCACCGCCCGCCCGCCAGGTGTCACTGAGCCCGGA 5105
 QY 486 CACAGGCGAGCCCGGGCTCCACCGCCCGCCCGCCAGGTGTCACTGAGCCCGGA 545
 Db 5106 CANNAGGCGANNCCGGGCTCCACCGCCCGCCCGCCAGGTGTCACTGAGCCCGGA 5165
 QY 546 CACAGGCGAGCCCGGGCTCCACCGCCCGCCCGCCAGGTGTCACTGAGCCCGGA 605
 Db 5166 CANNAGGCGANNCCGGGCTCCACCGCCCGCCCGCCAGGTGTCACTGAGCCCGGA 5225
 QY 606 CAACAGGCGAGCCCGGGCTCCACCGCCCGCCCGCCAGGTGTCACTGAGCCCGGA 665
 Db 5226 CANNAGGCGANNNTTGGGCTCCACCGCCCGCCCGCCAGGTGTCACTGAGCCCGGA 5285
 QY 666 CTCTGATCAGGCTCAGCTTCTACTCTGTGTCAGAAAGGACCTGTGCAAGGCTACAC 725
 Db 5286 CTCTGATCAGGCTCAGCTTCTACTCTGTGTCAGAAAGGACCTGTGCAAGGCTACAC 5345
 QY 726 AACCCAGCAGAGAGACCTCCACCGAGATTCCAGCCACCACTGTACTCTTAC 785
 Db 5346 AACCCAGCAGAGAGACCTCCACCGAGATTCCAGCCACCACTGTACTCTTAC 5405
 QY 786 CACCTTGGCAGCAGTACGACCAAGACTGATGCAAGTACACTCACTACAGTAC 845
 Db 5406 CACCTTGGCAGCAGTACGACCAAGACTGATGCAAGTACACTCACTACAGTAC 5465
 QY 846 TCCTCTCACTCTCTCCATACAGACATTTCTCCAGTTGTCTACTGGGGTCTTTCTT 905
 Db 5466 TCCTCTCACTCTCTCCATACAGACATTTCTCCAGTTGTCTACTGGGGTCTTTCTT 5525
 QY 906 TTTCTGTCTTTTCAATTCAATTAACCTCAGATTAAATTCCTCTGAGAAATCCAGAC 965
 Db 5526 TTTCTGTCTTTTCAATTCAATTAACCTCAGATTAAATTCCTCTGAGAAATCCAGAC 5585
 QY 966 CGACTTACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTGCAATTTATAACA 1025
 Db 5586 CGACTTACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTGCAATTTATAACA 5645
 QY 1026 AGGGGGTTTTCTGGGCTCTCCATATTAATGTTAGGCGCAGATCTGAGTACAT 1085
 Db 5646 AGGGGGTTTTCTGGGCTCTCCATATTAATGTTAGGCGCAGATCTGAGTACAT 5705
 QY 1086 GACTCTGGCTTCCGAGAGTACATCATATGTCACAGCTGAGACACAGTTCAATCA 1145
 Db 5706 GACTCTGGCTTCCGAGAGTACATCATATGTCACAGCTGAGACACAGTTCAATCA 5765
 QY 1146 GTATTAACCGAAGCAGCTCTCGATATTAACCTGACATCTCAGAGTCAAGTGAATCA 1205
 Db 5766 GTATTAACCGAAGCAGCTCTCGATATTAACCTGACATCTCAGAGTCAAGTGAATCA 5825

QY	1206	TGTGCAATTTCTTTTCTCTGTGCCAATCTTG3636CTGG36GTGCGAGCTTG36GATCGCGCT	1255
Db	5826	TGTGCAATTTCTTTTCTCTGTGCCAATCTTG3636CTGG36GTGCGAGCTTG36GATCGCGCT	5885
QY	1266	GGTGGGTCTGGTCTGTGTTCTGGTTGGTCCCTGGACATGTCTATCTCATTTGGCTTGGCTGT	1325
Db	5886	GGTGGGTCTGGTCTGTGTTCTGGTTGGTCCCTGGACATGTCTATCTCATTTGGCTTGGCTGT	5945
QY	1326	CTGTCAATGTCGGCCGGAAGAACTAAGGGGACAGTGGACATCTTTCCAGCCGG36ATACCTA	1385
Db	5946	CTGTCAATGTCGGCCGGAAGAACTAAGGGGACAGTGGACATCTTTCCAGCCGG36ATACCTA	6005
QY	1386	CCATCTTATGAGCAGATACCCACCTACACACACCATG36GCTATGTGCCCCCTAGCAG	1445
Db	6006	CCATCTTATGAGCAGATACCCACCTACACACACCATG36GCTATGTGCCCCCTAGCAG	6065
QY	1446	TACCATGTGTAGCCCCCTATATGAAAGGTTTCTGACAGTAATGTGTGACAGACCTCTTAA	1505
Db	6066	TACCATGTGTAGCCCCCTATATGAAAGGTTTCTGACAGTAATGTGTGACAGACCTCTTAA	6125
QY	1506	CACAAACCCAGAGTGGCAGACCACTTCTGCCAATTGTAG3636CACTGTG36CTCTAGAC	1565
Db	6126	CACAAACCCAGAGTGGCAGACCACTTCTGCCAATTGTAG3636CACTGTG36CTCTAGAC	6185
QY	1566	TGAGTGG 1572	
Db	6186	TGAGTGG 6192	
RESULT 13			
LOCUS	AX334899	4139 bp	DNA linear PAT 09-JAN-2002
DEFINITION	Sequence 5408 from Patent WO0194629.		
ACCESSION	AX334899		
VERSION	AX334899.1	GI:18125618	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 Young, P.E., Augustus, M., Carter, K.C., Ehner, R., Endress, G., Horrigan, S., Soppet, D.R. and Weaver, Z.		
TITLE	Cancer gene determination and therapeutic screening using signature gene sets		
JOURNAL	Patent: WO 0194629-A 5408 13-DEC-2001; Avalon Pharmaceuticals (US)		
FEATURES	Location/Qualifiers		
source	1..4139 /organism="Homo sapiens" /db_xref="taxon:9606"		
BASE COUNT	632 a 1910 c 1055 g 542 t		
ORIGIN			
Query Match	75.7%; Score 1190.2; DB 6; Length 4139;		
Best Local Similarity	94.8%; Pred. No. 2.2e-207;		
Matches 1231; Conservative	0; Mismatches 68; Indels 0; Gaps 0;		
QY	249	CAGCCCGGTTACAGCTCTCCACCACTACAGG9ACAGGATATCTCTG36CCG36CCAC	308
Db	2545	CAGCCCGGTTACAGCTCTCCACCGCCCCCAGCCACAGGTATCACTCG6CCCG36AC	2604
QY	309	GGAAACAGTTTACAGTTACGTGCCACTTG36G9ACAGATGTCACTTCGGTCCCAATCA	368
Db	2605	CAGGCTCGGCCCCCG36CTCCACCGCCCCCAGCCACAGGTATCACTTCG36CCCG36AC	2664
QY	369	CAGGCGAGCCCTGG36CTCCACCACTCCGCGACCAACAGATATCACTCAGCCCG36CAA	428
Db	2665	CAGGCTCGGCCCCCG36CTCCACCGCCCCCAGCCACAGGTATCACTCG36CCCG36AC	2724
QY	429	CAGGCGAGCCCTGG36CTCCACCGCCCCCAGCCACAGGTATCACTTCG36CCCG36AC	488
Db	2725	CAGGCTCGGCCCCCG36CTCCACCGCCCCCAGCCACAGGTATCACTTCG36CCCG36AC	2784

QY	489	CAGGCGCCCGGGGCTTCAACGCCCCCGCCGCCACAGGTGTCACTTGCGCCCGGACAA	548
Db	2785	CAGGCGCGGCCCCGGGGCTTCCACGCCCCCGCCAGCCACAGGTGTCACTTGCGCCCGGACAA	2844
QY	549	CAGGCGCGGCCCCGGGGCTTCCACGCCCCCGCCAGCCACAGGTGTCACTTGCGCCCGGACAA	608
Db	2845	CAGGCGCGGCCCCGGGGCTTCCACGCCCCCGCCAGCCACAGGTGTCACTTGCGCCCGGACAA	2904
QY	609	CAGGCGGCGCTTGGGGCTTCCACGCCCCCTTCAGTGTCAAAATGTCACTTGCGGCTCAAGGCTC	668
Db	2905	CAGGCGGCGCTTGGGGCTTCCACGCCCCCTTCAGTGTCAAAATGTCACTTGCGGCTCAAGGCTC	2964
QY	669	TGCATCAGGCTCAGCTTCTACTGTGGTGAACAAGGACCTCTGCCAGGGCTTACCAAC	728
Db	2965	TGCATCAGGCTCAGCTTCTACTGTGGTGAACAAGGACCTCTGCCAGGGCTTACCAAC	3024
QY	729	CCCAAGCAGCAAGAGACCTCCACCAGACTTCCAGGCAACAACACTGTAACTCCACAC	788
Db	3025	CCCAAGCAGCAAGAGACCTCCACCAGACTTCCAGGCAACAACACTGTAACTCCACAC	3084
QY	789	CCTTGCAGCATAGACCAAGAAGTAGCCAGTAGACCTCACATATGACGGTACTCC	848
Db	3085	CCTTGCAGCATAGACCAAGAAGTAGCCAGTAGACCTCACATATGACGGTACTCC	3144
QY	849	TCTCACCCTTCCATACAGACCTTCCGCCAGTGTCTACTGGGGTCTCTTTCTTTT	908
Db	3145	TCTCACCCTTCTCATATCAGACCTTCCGCCAGTGTCTACTGGGGTCTCTTTCTTTT	3204
QY	909	CCTGTCTTTTCAATTTCAAACTCCAGTTTAATCTCTCTGGAAGATCCAGCACGA	968
Db	3205	CCTGTCTTTTCAATTTCAAACTCCAGTTTAATCTCTCTGGAAGATCCAGCACGA	3264
QY	969	CTACTACAAAGGCTGCGAGAGACATTTCTGAAATGTTTTGAGATTTTAAACAAG	1028
Db	3265	CTACTACAAAGGCTGCGAGAGACATTTCTGAAATGTTTTGAGATTTTAAACAAG	3324
QY	1029	GGGTTTTCTGGGCGCTTCCAAATTAAGTTCAAGGCGAGATCTGTGGTGTCAATTTAC	1088
Db	3325	GGGTTTTCTGGGCGCTTCCAAATTAAGTTCAAGGCGAGATCTGTGGTGTCAATTTAC	3384
QY	1089	TCTGGGCTTCCGAGAAGGTACATCAATGTCCACAGCTGGAGACACAGTTCAATCAGTA	1148
Db	3385	TCTGGGCTTCCGAGAAGGTACATCAATGTCCACAGCTGGAGACACAGTTCAATCAGTA	3444
QY	1149	TAAACCGGAAGCAGCTCTCGAATTAACCTGACATCTCAGACGTCAAGGTGAATCAGT	1208
Db	3445	TAAACCGGAAGCAGCTCTCGAATTAACCTGACATCTCAGACGTCAAGGTGAATCAGT	3504
QY	1209	GCGATTTCTTTCTCGCGCCAGCTGTGGGGCTGGGGGTGCGCAGGCTGGGGACCTGCGTGT	1268
Db	3505	GCGATTTCTTTCTCGCGCCAGCTGTGGGGCTGGGGGTGCGCAGGCTGGGGACCTGCGTGT	3564
QY	1269	GATGCTGTCTGTGTTCTGTGGTGGCTGGCATTTCTATCTCATTTGCGCTGTCTG	1328
Db	3565	GATGCTGTCTGTGTTCTGTGGTGGCTGGCATTTCTATCTCATTTGCGCTGTCTG	3624
QY	1329	TCAAGTGCAGCCAAAGAACTACAGGGGAGCTGACACTTTTCCAGCCGGGANTACTTACA	1388
Db	3625	TCAAGTGCAGCCAAAGAACTACAGGGGAGCTGACACTTTTCCAGCCGGGANTACTTACA	3684
QY	1389	TCTATAGCGAATACCCCACTTACACACCCATGGGGCTATATGGCCCCCTTAGCAATAC	1448
Db	3685	TCTATATAGGAGATACCCCACTTACACACCCATGGGGCTATATGGCCCCCTTAGCAATAC	3744
QY	1449	CGATGTAGCCCTTATAGAAAGTTTTCGACGATTAATGTGTGACAGACCTCTTTTAC	1508
Db	3745	CGATGTATAGCCCTTATAGAAAGTTTTCGACGATTAACGCTGTGACAGACCTCTTTTAC	3804
QY	1509	AAACCCAGACGTGGACCACTTCTGCCAATCTTTAGGG	1547
Db	3805	AAACCCAGACGTGGACCGCGCTTCTGCCAATCTTTAGGG	3843

RESULT 14
 AX335372 4139 bp DNA linear PAT 09-JAN-2002
 LOCUS Sequence 5881 from Patent WO0194629.
 DEFINITION AX335372
 ACCESSION AX335372
 VERSION AX335372.1 GI:1812691
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endress, G.,
 Horigan, S., Soppet, D. R. and Weaver, Z.
 TITLE Cancer gene determination and therapeutic screening using signature
 gene sets
 JOURNAL Patent: WO 0194629-A 5881 13-DEC-2001;
 Avalon Pharmaceuticals (US)
 FEATURES
 source Location/Qualifiers
 1. 4139
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 632 a 1910 c 1055 g 542 t
 ORIGIN
 Query Match 75.7%; Score 1190.2; DB 6; Length 4139;
 Best Local Similarity 94.8%; Pred. No. 2.2e-207;
 Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 909 CTTGCTTTTCAATTCAAACTTCAGTTTAAATTCCTCTGGAAGATCCACGCGA 968
 Db 3205 CTTGCTTTTCAATTCAAACTTCAGTTTAAATTCCTCTGGAAGATCCACGCGA 3264
 QY 969 CTATACCAAGAGCTGCAAGAGAGCAATTTGGAATGTTTGGAGATTTAACAAG 1028
 Db 3265 CTATACCAAGAGCTGCAAGAGAGCAATTTGGAATGTTTGGAGATTTAACAAG 3324
 QY 1029 GGGTTTTCTGGGCTCTCCAAATATTAGTTGAGCCAGATCTGTGTGATCAATTGAC 1088
 Db 3325 GGGTTTTCTGGGCTCTCCAAATATTAGTTGAGCCAGATCTGTGTGATCAATTGAC 3384
 QY 1089 TCTGGCTTCCGAGAGATGATCAATATGTCACGAGCTGAGACACAGTTCAATCAGTA 1148
 Db 3385 TCTGGCTTCCGAGAGATGATCAATATGTCACGAGCTGAGACACAGTTCAATCAGTA 3444
 QY 1149 TAAACGGAAGAGCTCTCCGATTAATACGATCTGAGACGTCAGGAGTCAATG 1208
 Db 3445 TAAACGGAAGAGCTCTCCGATTAATACGATCTGAGACGTCAGGAGTCAATG 3504
 QY 1209 GGCATTTCTTTTCTGCCAGTCTGAGCTGAGGCTGAGGCTGAGGCTGAGCTGCT 1268
 Db 3505 GGCATTTCTTTTCTGCCAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGCTGCT 3564
 QY 1269 GGTGCTGTCTGTGTTCTGTGTTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1328
 Db 3565 GGTGCTGTCTGTGTTCTGTGTTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 3624
 QY 1329 TCAAGTCCGCGAAGAACTACGAGGAGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1388
 Db 3625 TCAAGTCCGCGAAGAACTACGAGGAGCTGAGGCTGAGGCTGAGGCTGAGGCTG 3684
 QY 1389 TCTATGAGGAGTACCCCACTACCAACCCATGAGGCTGATGTCCTTACAGTAC 1448
 Db 3685 TCTATGAGGAGTACCCCACTACCAACCCATGAGGCTGATGTCCTTACAGTAC 3744
 QY 1449 GATGTGAGGCTCTATGAGAAAGTTTCTGAGGATTAATGAGGAGGCTGAGGCTG 1508
 Db 3745 GATGTGAGGCTCTATGAGAAAGTTTCTGAGGATTAATGAGGAGGCTGAGGCTG 3804
 QY 1509 AAACCCAGAGTGGAGGCACTTGTGCAACTTTGTAGGG 1547
 Db 3805 AAACCCAGAGTGGAGGCACTTGTGCAACTTTGTAGGG 3843

RESULT 15
 AX336712 4139 bp DNA linear PAT 09-JAN-2002
 LOCUS Sequence 7221 from Patent WO0194629.
 DEFINITION AX336712
 ACCESSION AX336712
 VERSION AX336712.1 GI:18127431
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endress, G.,
 Horigan, S., Soppet, D. R. and Weaver, Z.
 TITLE Cancer gene determination and therapeutic screening using signature
 gene sets
 JOURNAL Patent: WO 0194629-A 7221 13-DEC-2001;
 Avalon Pharmaceuticals (US)
 FEATURES
 source Location/Qualifiers
 1. 4139
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 632 a 1910 c 1055 g 542 t
 ORIGIN
 Query Match 75.7%; Score 1190.2; DB 6; Length 4139;
 Best Local Similarity 94.8%; Pred. No. 2.2e-207;

Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

OY 249 CAGCCCCGGTTGAGGCTCTCTCCACCACTCAGGAGAGATGACCTGTGGCCCGGAGCAC 308
Db 2545 CAGGCCCGGCGCGGCTCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGACAC 2604
OY 309 GGAACCAAGCTTCAAGTTCAAGCTGACCACTGGGGGAGACAGATGTGACCTGGTCCAGTAC 368
Db 2605 CAGGCCCGGCGCGGCTCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGACAC 2664
OY 369 CAGGCCCGGCGGCTCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGACAA 428
Db 2665 CAGGCCCGGCGGCTCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGACAC 2724
OY 429 CAAAGCCAGCCCGGAGCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGACAC 488
Db 2725 CAGGCCCGGCGGCTCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGACAC 2784
OY 489 CAGGCCCGGCGGCTCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGACAC 548
Db 2785 CAGGCCCGGCGGCTCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGACAC 2844
OY 549 CAGGCCCGGCGGCTCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGACAA 608
Db 2845 CAGGCCCGGCGGCTCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGACAA 2904
OY 609 CAGGCCCGGCGGCTCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGAGCTC 668
Db 2905 CAGGCCCGGCGGCTCTCCACCGCCCCCGAGCCAGCGGTTCACCTGGGCCCCGGAGCTC 2964
OY 669 TGCATCAGGCTCAGCTTCTAATCTGTGTGACAAAGGACCTCTGCCAGGGCTTACGACAAAC 728
Db 2965 TGCATCAGGCTCAGCTTCTAATCTGTGTGACAAAGGACCTCTGCCAGGGCTTACGACAAAC 3024
OY 729 CCCAGCCAGCAAGAGCACTCCACCAGCTTCCAGCCAGCACTGTATCTCTTACCAAC 788
Db 3025 CCCAGCCAGCAAGAGCACTCCACCAGCTTCCAGCCAGCACTGTATCTCTTACCAAC 3084
OY 789 CCTTCCAGCCATAGCAACCAAGACTGATCCAGTAGCACTGACCATATAGCAAGGTAACCTCC 848
Db 3085 CCTTCCAGCCATAGCAACCAAGACTGATCCAGTAGCACTGACCATATAGCAAGGTAACCTCC 3144
OY 849 TCTGACCTCTCCATCAGACAGCACTTCTCCAGTGTCTACTGGGGTCTCTTCTTTT 908
Db 3145 TCTGACCTCTCCATCAGACAGCACTTCTCCAGTGTCTACTGGGGTCTCTTCTTTT 3204
OY 909 CCTGTCTTTTCACTTTCAAACTCCAGTTTAACTCTCTGGAAGATCCGAGCACCGA 968
Db 3205 CCTGTCTTTTCACTTTCAAACTCCAGTTTAACTCTCTGGAAGATCCGAGCACCGA 3264
OY 969 CTACTACCAAGAGCTGAGAGAGACATTTCTGAAATGTTTTGCAATTATTAACAAG 1028
Db 3265 CTACTACCAAGAGCTGAGAGAGACATTTCTGAAATGTTTTGCAATTATTAACAAG 3324
OY 1029 GGGTTTTTCTGGGCTCTTCCAAATATTAAGTTCAAGGCCAGGATCTGTGGTGAATTAAC 1088
Db 3325 GGGTTTTTCTGGGCTCTTCCAAATATTAAGTTCAAGGCCAGGATCTGTGGTGAATTAAC 3384
OY 1089 TCTGGGCTTCCGAGAAAGGTACCATCAATGTCCAGACGTGAGACACAAGTTCAATCAATA 1148
Db 3385 TCTGGGCTTCCGAGAAAGGTACCATCAATGTCCAGACGTGAGACACAAGTTCAATCAATA 3444
OY 1149 TAAACAGGAAGAGAGCTCTGATTAATCAATCAAGATCTGAGACGTCAAGCTGAGTCAAT 1208
Db 3445 TAAACAGGAAGAGAGCTCTGATTAATCAATCAAGATCTGAGACGTCAAGCTGAGTCAAT 3504
OY 1209 GCCATTTCTCTTCTCTGCCAGATCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCT 1268
Db 3505 GCCATTTCTCTTCTCTGCCAGATCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCT 3564
OY 1269 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1328
Db 3565 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3624

```

```

OY 1329 TCAGTCCCGCCGAAAGATTAACGGGCACTGGAATTTTTCAGCCCGGAGTACCTTACCA 1388
Db 3625 TCAGTCCCGCCGAAAGATTAACGGGCACTGGAATTTTTCAGCCCGGAGTACCTTACCA 3684
OY 1389 TCCATAGAGAGATACCCCACTTACCAACCCATGAGCGCTATGTGCCCCCTAGCAGTAC 1448
Db 3685 TCCATAGAGAGATACCCCACTTACCAACCCATGAGCGCTATGTGCCCCCTAGCAGTAC 3744
OY 1449 CGATGTAGCCCTTATGAAAGTTTTCAGAGTAAATGTGAGGAGCAGGCTCTTACAC 1508
Db 3745 CGATGTAGCCCTTATGAAAGTTTTCAGAGTAAATGTGAGGAGCAGGCTCTTACAC 3804
OY 1509 AAACCCAGAGTGGCAGCCACTTGTGCCACTGTAGGG 1547
Db 3805 AAACCCAGAGTGGCAGCCACTTGTGCCACTGTAGGG 3843

```

Search completed: May 8, 2003, 16:17:28
Job time : 4420 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 22:59:10 ; Search time 1.16071 Seconds

(without alignments)
1033.205 Million cell updates/sec

Title: US-09-658-621B-26

Perfect score: 44

Sequence: 1 ISEMFLOIY 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	9	22	AAU00563
2	44	100.0	10	22	AAU00566
3	44	100.0	30	23	AAU84995
4	44	100.0	216	21	AAU71022
5	44	100.0	239	17	AAU89423
6	44	100.0	240	17	AAU89422
7	44	100.0	255	17	AAU89420
8	44	100.0	264	17	AAU89421
9	44	100.0	273	17	AAU89418
10	44	100.0	282	17	AAU89419

11	44	100.0	312	23	AAU84810	Human MUC1R consen
12	44	100.0	321	21	AAU84346	Human cancer assoc
13	44	100.0	327	16	AAU89628	Glycoprotein 39 C
14	44	100.0	338	21	AAU71028	Ubiquitin-E. coli
15	44	100.0	348	13	AAU27662	C-terminal region
16	44	100.0	455	13	AAU23973	Transmembrane form
17	44	100.0	455	21	AAU71024	Human Mucin 1 (MUC
18	44	100.0	473	22	AAU09508	Human mucin-1 (MUC
19	44	100.0	475	22	AAU00573	Human MUC1 polypep
20	44	100.0	475	23	AAU77476	Human MUC1 seq ID
21	44	100.0	495	22	AAU00539	Human MUC1 polypep
22	44	100.0	577	21	AAU71030	Ubiquitin-E. coli
23	44	100.0	5546	23	AAU85008	Human melanocyte d
24	41	93.2	508	19	AAU77233	Human melanocyte d
25	36	81.8	1091	20	AAU30337	Human melanocyte d
26	33	75.0	57	21	AAU60595	Protein encoded by
27	33	75.0	61	21	AAU60354	Arabidopsis thalia
28	33	75.0	71	21	AAU47921	Arabidopsis thalia
29	33	75.0	1005	21	AAU48223	Arabidopsis thalia
30	33	75.0	1057	21	AAU48222	Arabidopsis thalia
31	33	75.0	1114	21	AAU48221	Arabidopsis thalia
32	32	72.7	303	13	AAU27654	Human calcium chan
33	32	72.7	1848	22	AAU58215	Drosophila melano
34	32	72.7	2262	19	AAU56737	Calcium ion channe
35	32	72.7	2265	22	AAU69072	Rabbit P/O-type ca
36	32	72.7	2266	16	AAU71008	Human neuronal cal
37	32	72.7	2266	21	AAU10580	Human calcium chan
38	32	72.7	2424	21	AAU78901	Calcium channel al
39	32	72.7	2510	16	AAU71007	Human neuronal cal
40	32	72.7	2510	21	AAU10579	Human calcium chan
41	31	70.5	242	21	AAU10579	Bacteriophage Dp-1
42	31	70.5	522	22	AAU46708	R. marinus bacteri
43	31	70.5	871	22	AAU46727	R. marinus bacteri
44	31	70.5	1336	22	AAU59655	Drosophila melano
45	31	70.5	1639	19	AAU54145	P. falciparum synt

ALIGNMENTS

RESULT 1	AAU00563	standard; Peptide; 9 AA.
ID	AAU00563	
AC	AAU00563	
XX		
DT	12-SEP-2001	(first entry)
XX		
DE	Human MUC1 polypeptide derivative #24.	
XX		
KW	Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-I;	
KW	glycoprotein; cytotoxic T lymphocytes; T cell response; cancer; vaccine;	
KW	cancer gene therapy; diagnosis; treatment; inflammatory disorder;	
KW	organ transplant rejection; graft versus host disease.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200118035-A2.	
XX		
PD	15-MAR-2001.	
XX		
PF	07-SEP-2000; 2000WO-EP08761.	
XX		
PR	08-SEP-1999; 99GB-0021242.	
PR	10-SEP-1999; 99EP-0402237.	
PR	03-MAR-2000; 2000US-0187215.	
XX		
PA	(TRGE) TRANSGENE SA.	
PA	(IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.	
XX		
PI	Taylor-Papadimitriou J, Heukamp LC, Offringa R, Melief CJM;	
PI	Acres B, Thomas M;	
XX		

DR WPI: 2001-235187/24.
DR N-PSDB: AAS00609.
XX New antigenic polypeptides of MUC-1 protein which activate cytotoxic T
PT lymphocyte proteins and their analogues, useful for identifying a major
PT histocompatibility complex class I restricted T cell response and for
PT diagnosing cancer -
XX
PS Claim 24, Page 74; 81pp; English.
XX The sequence represents a human MUC1 polypeptide derivative. Derivative
CC antigenic peptides of MUC1 protein bind at least one major
CC histocompatibility complex class I (MHC-I) glycoprotein, which activates
CC cytotoxic T lymphocytes to induce a protective response against tumors.
CC Diagnosis of cancer involves determining the presence or absence in a
CC host cell of MHC class I restricted T cell response to a MUC1 derivative,
CC where the presence of the MHC class I restricted T cell response
CC indicates that the host has cancer. Measurement of the level of MHC class
CC I restricted T cell response is also useful to monitor the severity of
CC cancer, a larger response indicating a more severe cancer. MUC1
CC derivatives are useful in cancer therapy and to follow MUC1 specific
CC immune responses in patients during the course of disease and/or
CC treatment. MUC1 DNA is useful in cancer gene therapy, vaccination and
CC diagnosis. Compositions of the sequences are used in vaccines and
CC treatments against cancer or diseases caused by an immune response, such
CC as an inflammatory disorder, organ transplant rejection or graft versus
CC host disease.
XX
SQ Sequence 9 AA;
XX
XX Query Match 100.0%; Score 44; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ISEMPLOIY 9
DB 1 ISEMPLOIY 9	
RESULT 2
AAU00566
ID AAU00566 standard; Peptide; 10 AA.
XX
AC AAU00566;
XX
DT 12-SEP-2001 (first entry)
XX
DE Human MUC1 polypeptide derivative #27.
XX
XX Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-I;
XX glycoprotein; cytotoxic T lymphocytes; T cell response; cancer; vaccine;
XX cancer gene therapy; diagnosis; treatment; inflammatory disorder;
XX organ transplant rejection; graft versus host disease.
XX
XX Homo sapiens.
OS
XX WO200118035-A2.
XX
XX 15-MAR-2001.
XX
XX 07-SEP-2000; 2000WO-EP08761.
XX
XX 08-SEP-1999; 99GB-0021242.
XX 10-SEP-1999; 99EP-0402237.
XX 03-MAR-2000; 2000US-0187215.
XX
XX (TRGE) TRANSGENE SA.
XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
XX
XX Taylor-Papadimitriou J, Heukamp IC, Offringa R, Melief CUM;
XX Acres B, Thomas M;
XX WPI: 2001-235187/24.
DR

DR N-PSDB: AAS00712.
XX
XX New antigenic polypeptides of MUC-1 protein which activate cytotoxic T
PT lymphocyte proteins and their analogues, useful for identifying a major
PT histocompatibility complex class I restricted T cell response and for
PT diagnosing cancer -
XX
PS Claim 2; Page 74; 81pp; English.
XX The sequence represents a human MUC1 polypeptide derivative. Derivative
CC antigenic peptides of MUC1 protein bind at least one major
CC histocompatibility complex class I (MHC-I) glycoprotein, which activates
CC cytotoxic T lymphocytes to induce a protective response against tumors.
CC Diagnosis of cancer involves determining the presence or absence in a
CC host cell of MHC class I restricted T cell response to a MUC1 derivative,
CC where the presence of the MHC class I restricted T cell response
CC indicates that the host has cancer. Measurement of the level of MHC class
CC I restricted T cell response is also useful to monitor the severity of
CC cancer, a larger response indicating a more severe cancer. MUC1
CC derivatives are useful in cancer therapy and to follow MUC1 specific
CC immune responses in patients during the course of disease and/or
CC treatment. MUC1 DNA is useful in cancer gene therapy, vaccination and
CC diagnosis. Compositions of the sequences are used in vaccines and
CC treatments against cancer or diseases caused by an immune response, such
CC as an inflammatory disorder, organ transplant rejection or graft versus
CC host disease.
XX
SQ Sequence 10 AA;
XX
XX Query Match 100.0%; Score 44; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ISEMPLOIY 9
DB 1 ISEMPLOIY 9	
RESULT 3
AAU84995
ID AAU84995 standard; Peptide; 30 AA.
XX
AC AAU84995;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human MUC1R segment 9.
XX
XX Svine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
XX viral infection; human immunodeficiency virus; melanoma;
XX bacterial infection; Salmonella; Legionella; parasitic infection;
XX Trypanosoma; Toxoplasma; Giardia.
XX
XX Homo sapiens.
OS
XX WO200190197-A1.
XX
XX 29-NOV-2001.
XX
XX 25-MAY-2001; 2001WO-AU00622.
XX
XX 26-MAY-2000; 2000AU-0007761.
XX
XX (AUSU) UNIV AUSTRALIAN NAT.
XX
XX Thomson SA, Ramshaw IA;
XX WPI: 2002-147575/19.
XX N-PSDB: ABK36815.
XX
XX New synthetic polypeptides having several different segments of at
PT least one parent polypeptide linked together differently compared to
PT the linkage in the parent polypeptide, for inducing immune response
PT

PT against a pathogen or cancer -
XX
XX Example 3; Fig 27; 364pp; English.

CC The invention relates to a new synthetic polypeptide (I) comprising
CC several different segments of at least one parent polypeptide linked
CC together in a different relationship relative to their linkage in the
CC parent polypeptide to impede, abrogate or otherwise alter at least one
CC function associated with the parent polypeptide and for inducing an
CC immune response against a pathogen or cancer. Also included are a
CC synthetic polynucleotide encoding and a computer system for
CC designing the synthetic polypeptides. The synthetic polypeptides and
CC polynucleotides are referred to as a Savine. The synthetic polypeptide is
CC useful for modulating immune responses preferably directed against a
CC pathogen or a cancer (e.g., cancers of the lung, breast, ovary, cervix,
CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
CC Compositions comprising the polypeptide may be used in the treatment or
CC prophylaxis against viral (such as infections caused by HIV (human
CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
CC Trypanosoma, Toxoplasma and Giardia) infections. The present
CC sequence is a peptide derived from a parent protein used to
CC construct a savine of the invention.

XX Sequence 30 AA;

Query Match 100.0%; Score 44; DB 23; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLOIY 9
| | | | |
DB 12 ISEMFLOIY 20

RESULT 4
AA71022
ID AA71022 standard; Protein; 216 AA.

XX AA71022;

XX 29-AUG-2000 (first entry)

DE Human Mucin 1 (MUC-1) protein fragment #3.

XX Human; Mucin 1; tumour; pMRS30 expression vector;
XX anti-tumour; therapy; immune response; cytostatic; vaccine.

XX Homo sapiens.

XX WO200025827-A2.

XX 11-MAY-2000.

XX 18-OCT-1999; 99WO-EP07874.

XX 30-OCT-1998; 98IT-M102330.

XX (MENA) MENARINI RIGERCHE SPA.

XX Parente D, Di Massimo AM, De Santis R;

XX WPI: 2000-365410/31.

XX N-PEDB; AAD00386.

PT Composition containing one or more DNA molecules encoding fragments of
PT a Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in
PT anti-tumor therapy -

PS Claim 16; Fig 3; 56pp; English.

XX The present sequence is a fragment of human Mucin 1 (MUC-1), an
XX antigenic protein overexpressed in tumour cells. The sequence was
XX obtained from BT20 tumour cells. The corresponding DNA sequence
XX is cloned into a pMRS30 expression vector and used in pharmaceutical
XX composition e.g. vaccine for inducing an antigen-specific anti-tumour
XX immune response. Composition containing this DNA molecule is useful in
XX anti-tumour therapy of patients affected with tumours characterised
XX by high MUC-1 expression.

XX Sequence 216 AA;

Query Match 100.0%; Score 44; DB 21; Length 216;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLOIY 9
| | | | |
DB 108 ISEMFLOIY 116

RESULT 5
AAR89423
ID AAR89423 standard; Protein; 239 AA.

XX AAR89423;

XX 24-APR-1996 (first entry)

DE Mucin-derived protein MUC1/V/alt.

XX MUC1; MUC1/X/alt; MUC1/Y/alt; MUC1/V/alt; MUC1/V/alt;

XX MUC1/W/alt; MUC1/Z/alt; mucin; breast cancer;

XX receptor; diagnosis; imaging; therapy.

XX Homo sapiens.

XX WO9603502-A2.

XX 08-FEB-1996.

XX 21-JUL-1995; 95WO-IB00627.

XX 26-JUL-1994; 94IL-0110464.

XX (UPRA-) UNIV RAMOT APPL RES & IND DEV LTD.

XX Wreschner DH;

XX WPI: 1996-117047/12.

XX N-PEDB; AAT10682.

XX Claim 6; Fig 6D; 79pp; English.

XX Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
XX MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly and
XX differentially expressed in human breast cancer tissue and significantly
XX enhance the in vivo tumorigenic potential of mammary epithelial cells.
XX They serve as cell surface receptor molecules participating in signal
XX transduction. The proteins can be obtd. by expression of encoding cDNA
XX (see AAT10677-82) in recombinant host cells. They are used in the
XX treatment of human breast cancer and as diagnostic reagents. Receptor
XX ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
XX isolated.

Query Match 100.0%; Score 44; DB 17; Length 239;

Best Local Similarity 100.0%; Pred. No. 0.34; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
Db 57 ISEMFLQIY 65

RESULT 6

AA89422

ID AAR89422 standard; Protein: 240 AA.

AC AAR89422;

DT 24-APR-1996 (first entry)

DE Mucin-derived protein MUC1/V.

XX MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
XX MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
XX receptor; diagnosis; imaging; therapy.

OS Homo sapiens.

PN WO9603502-A2.

PD 08-FEB-1996.

PE 21-JUL-1995; 95WO-IB00627.

PR 26-JUL-1994; 94IL-0110464.

PA (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.

PI Wreschner DH;

DR WPI; 1996-117047/12.

DR N-PSDB; AAT10681.

PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
PT cancer

PS Claim 6; Fig 6C; 79pp; English.

XX Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
XX MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly and
XX differentially expressed in human breast cancer tissue and significantly
XX enhance the in vivo tumorigenic potential of mammary epithelial cells.
XX They serve as cell surface receptor molecules participating in signal
XX transduction. The proteins can be obtd. by expression of encoding cDNA
XX (see AAT10677-82) in recombinant host cells. They are used in the
XX treatment of human breast cancer and as diagnostic reagents. Receptor
XX ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
XX isolated.

SQ Sequence 240 AA;

Query Match 100.0%; Score 44; DB 17; Length 240;

Best Local Similarity 100.0%; Pred. No. 0.34; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
Db 58 ISEMFLQIY 66

RESULT 7

ID AAR89420 standard; Protein: 255 AA.

AC AAR89420;

XX

DT 24-APR-1996 (first entry)

DE Mucin-derived protein MUC1/Y.

XX MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
XX MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
XX receptor; diagnosis; imaging; therapy.

OS Homo sapiens.

PN WO9603502-A2.

PD 08-FEB-1996.

PE 21-JUL-1995; 95WO-IB00627.

PR 26-JUL-1994; 94IL-0110464.

PA (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.

PI Wreschner DH;

DR WPI; 1996-117047/12.

DR N-PSDB; AAT10679.

PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
PT cancer

PS Claim 5; Fig 6A; 79pp; English.

XX Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
XX MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly and
XX differentially expressed in human breast cancer tissue and significantly
XX enhance the in vivo tumorigenic potential of mammary epithelial cells.
XX They serve as cell surface receptor molecules participating in signal
XX transduction. The proteins can be obtd. by expression of encoding cDNA
XX (see AAT10677-82) in recombinant host cells. They are used in the
XX treatment of human breast cancer and as diagnostic reagents. Receptor
XX ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
XX isolated.

SQ Sequence 255 AA;

Query Match 100.0%; Score 44; DB 17; Length 255;

Best Local Similarity 100.0%; Pred. No. 0.36; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
Db 73 ISEMFLQIY 81

RESULT 8

ID AAR89421 standard; Protein: 264 AA.

AC AAR89421;

DT 24-APR-1996 (first entry)

DE Mucin-derived protein MUC1/Y/alt.

XX MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
XX MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
XX receptor; diagnosis; imaging; therapy.

OS Homo sapiens.

PN WO9603502-A2.

PD 08-FEB-1996.

PF 21-JUL-1995; 95WO-IB00627.
 XX
 PR 26-JUL-1994; 94IL-0110464.
 XX
 PA (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.
 XX
 PI Wreschner DH;
 XX
 DR MPI: 1996-117047/12.
 XX
 DR N-PSDB; AAT10680.
 XX
 PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
 PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
 PT cancer
 XX
 PS Claim 5; Fig 6B; 79pp; English.
 XX
 CC Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
 CC MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly and
 CC differentially expressed in human breast cancer tissue and significantly
 CC enhance the in vivo tumorigenic potential of mammary epithelial cells.
 CC They serve as cell surface receptor molecules participating in signal
 CC transduction. The proteins can be obtd. by expression of encoding cDNA
 CC (see AAT10677-82) in recombinant host cells. They are used in the
 CC treatment of human breast cancer and as diagnostic reagents. Receptor
 CC ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
 CC isolated.
 CC
 CC Sequence 264 AA;
 SQ
 Query Match 100.0%; Score 44; DB 17; Length 264;
 Best Local Similarity 100.0%; Pred. No. 0.37; 0; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ISEMPLOITY 9
 DB 82 ISEMPLOITY 90
 XX
 RESULT 9
 AAR89418
 ID AAR89418 standard; Protein; 273 AA.
 XX
 AC AAR89418;
 XX
 DT 24-APR-1996 (first entry)
 XX
 DE Mucin-derived protein MUC1/X.
 XX
 KM MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
 KM MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
 KM receptor; diagnosis; imaging; therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO9603502-A2.
 XX
 PD 08-FEB-1996.
 XX
 PF 21-JUL-1995; 95WO-IB00627.
 XX
 PR 26-JUL-1994; 94IL-0110464.
 XX
 PA (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.
 XX
 PI Wreschner DH;
 XX
 DR MPI: 1996-117047/12.
 XX
 DR N-PSDB; AAT10677.
 XX
 PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
 PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
 PT cancer

XX
 PS Claim 4; Fig 5A; 79pp; English.
 XX
 CC Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
 CC MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly and
 CC differentially expressed in human breast cancer tissue and significantly
 CC enhance the in vivo tumorigenic potential of mammary epithelial cells.
 CC They serve as cell surface receptor molecules participating in signal
 CC transduction. The proteins can be obtd. by expression of encoding cDNA
 CC (see AAT10677-82) in recombinant host cells. They are used in the
 CC treatment of human breast cancer and as diagnostic reagents. Receptor
 CC ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
 CC isolated.
 CC
 CC Sequence 273 AA;
 SQ
 Query Match 100.0%; Score 44; DB 17; Length 273;
 Best Local Similarity 100.0%; Pred. No. 0.39; 0; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ISEMPLOITY 9
 DB 91 ISEMPLOITY 99
 XX
 RESULT 10
 AAR89419
 ID AAR89419 standard; Protein; 282 AA.
 XX
 AC AAR89419;
 XX
 DT 24-APR-1996 (first entry)
 XX
 DE Mucin-derived protein MUC1/X/alt.
 XX
 KM MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
 KM MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
 KM receptor; diagnosis; imaging; therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO9603502-A2.
 XX
 PD 08-FEB-1996.
 XX
 PF 21-JUL-1995; 95WO-IB00627.
 XX
 PR 26-JUL-1994; 94IL-0110464.
 XX
 PA (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.
 XX
 PI Wreschner DH;
 XX
 DR MPI: 1996-117047/12.
 XX
 DR N-PSDB; AAT10678.
 XX
 PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
 PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
 PT cancer
 XX
 PS Claim 4; Fig 5B; 79pp; English.
 XX
 CC Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
 CC MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly and
 CC differentially expressed in human breast cancer tissue and significantly
 CC enhance the in vivo tumorigenic potential of mammary epithelial cells.
 CC They serve as cell surface receptor molecules participating in signal
 CC transduction. The proteins can be obtd. by expression of encoding cDNA
 CC (see AAT10677-82) in recombinant host cells. They are used in the
 CC treatment of human breast cancer and as diagnostic reagents. Receptor
 CC ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
 CC isolated.

SQ Sequence 282 AA;
 Query Match 100.0%; Score 44; DB 17; Length 282;
 Best Local Similarity 100.0%; Pred. No. 0.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ISEMFLOIY 9
 |||||
 Db 100 ISEMFLOIT 108

RESULT 11

AAU84810

ID AAU84810 standard; Protein; 312 AA.

AC AAU84810;

DT 08-MAY-2002 (first entry)

DE Human MUC1R consensus sequence.

KW Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;

KW viral infection; human immunodeficiency virus; melanoma;

KW bacterial infection; Salmonella; Legionella; parasitic infection;

KW Trypanosoma; Toxoplasma; Giardia.

OS Homo sapiens.

XX WO200190197-A1.

XX 29-NOV-2001.

XX 25-MAY-2001; 2001WO-AU00622.

XX 26-MAY-2000; 2000AU-0007761.

XX (AUSU) UNIV AUSTRALIAN NAT.

PI Thomson SA, Ramshaw IA;

XX MPI; 2002-147575/19.

XX New synthetic polypeptides having several different segments of at
 PT least one parent polypeptide linked together differently compared to
 PT the linkage in the parent polypeptide, for inducing immune response
 PT against a pathogen or cancer

PS Example 3; Fig 27; 364pp; English.

XX The invention relates to a new synthetic polypeptide (I) comprising
 CC several different segments of at least one parent polypeptide linked
 CC together in a different relationship relative to their linkage in the
 CC parent polypeptide to impede, abrogate or otherwise alter at least one
 CC function associated with the parent polypeptide and for inducing an
 CC immune response against a pathogen or cancer. Also included are a
 CC synthetic polynucleotide encoding and a computer system for
 CC designing the synthetic polypeptides. The synthetic polypeptides and
 CC polynucleotides are referred to as a Savine. The synthetic polypeptide is
 CC useful for modulating immune responses preferably directed against a
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
 CC Salmonella, Streptococcal, Legionella and Mycobacterium) or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present
 CC sequence is a consensus sequence for a parent protein used to design a
 CC Savine of the invention.

SQ Sequence 312 AA;
 Query Match 100.0%; Score 44; DB 23; Length 312;
 Best Local Similarity 100.0%; Pred. No. 0.45;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ISEMFLOIY 9
 |||||
 Db 130 ISEMFLOIT 138

RESULT 12

AAB43416

ID AAB43416 standard; Protein; 321 AA.

AC AAB43416;

DT 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:861.

KW Human; cancer associated gene; cancer antigen; detection; cancer;

KW diagnosis; cytostatic; proliferative; vunerary; immunomodulator;

KW antidiabetic; antiaesthetic; antineumatic; antiarthritic; antiviral;

KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;

KW dermatological; neuroprotective; thrombolytic; coagulant; nocrotic;

KW vasotropic; antipostatic; antiangiogenic; gene therapy; inflammation;

KW immune disorder; haematopoietic cell disorder; autoimmune disorder;

KW allergic reaction; graft versus host disease; organ rejection;

KW haemostatic; thrombolytic; cardiovascular disorder; infection;

KW neurological disease; drug screening.

XX Homo sapiens.

XX WO200055350-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05882.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX MPI; 2000-58753/55.

XX N-PSDB; AAC77625.

PS Claim 11; Page 1414-1416; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given
 CC in AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vunerary; immunomodulator;
 CC antidiabetic; antiaesthetic; antineumatic; antiarthritic;
 CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
 CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
 CC nocrotic; vasotropic; antipostatic and antiangiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to

CC AACT8457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 321 AA;

Query Match 100.0%; Score 44; DB 21; Length 321;
 Best Local Similarity 100.0%; Pred. No. 0.46;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMFLQIY 9
 |||||
 DB 139 ISEMFLQIY 147

RESULT 13
 AAR6298 *
 ID AAR6298 standard; Peptide; 327 AA.
 XX
 AC AAR6298;

DT 26-JUL-1996 (first entry)

DE Glycoprotein 39 C terminal fragment.

XX Glycoprotein 39; gp39; lambda g11 cDNA library; gastric cancer;
 KW cell line KATO-III; tumour; immune abnormality; marker;
 KW inflammatory disease.

OS Homo sapiens.

PN JP07051065-A.

PD 28-FEB-1995.

PF 21-FEB-1992; 92JP-0035085.

PR 21-FEB-1992; 92JP-0035085.

PA (NIKO-) NIPPON KOTAI KENKYUSHO KK.
 XX (UYKA-) UNIV KAGOSHIMA.

DR WPI; 1995-127356/17.

DR N-PSDB; AAT29190.

PT Glyco:protein 39 gene - used in the mass production of glyco:protein
 PT 39, for use as tumour and immune abnormality marker

PS Claim 2; Page 9-10; 10pp; Japanese.

CC The sequences given in AAR6297-98 represent portions of glycoprotein
 CC 39. This sequence represents the C terminal portion of the glycoprotein
 CC 39 gene. The cDNA encoding this sequence was isolated from lambda g11
 CC cDNA library derived from the gastric cancer cell line KATO-III.
 CC Glycoprotein 39 is expected to be used as a tumour marker, an immune
 CC abnormality marker or a marker for various inflammatory diseases.

SQ Sequence 327 AA;

Query Match 100.0%; Score 44; DB 16; Length 327;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMFLQIY 9
 |||||
 DB 145 ISEMFLQIY 153

RESULT 14
 AAY71028
 ID AAY71028 standard; Protein; 338 AA.
 XX
 AC AAY71028;

DT 29-AUG-2000 (first entry)

XX Ubiquitin-E. coli LacI-human Mucin 1 fusion protein #3.

XX Ubiquitin; LacI; beta-galactosidase; fusion protein;

XX human; Mucin 1; MUC-1; tumour; pMR30 expression vector;

XX anti-tumour; therapy; immune response; cytostatic; vaccine.

OS Chimeric - Homo sapiens.

XX Chimeric - Escherichia coli.

XX Key Location/Qualifiers

FT Region 1..123
 FT /label= UBILacI protein
 FT /note= "contains ubiquitin sequence fused to a
 FT portion of E. coli LacI"
 FT 124..338
 FT /label= Human_MUC-1_fragment

XX WO200025827-A2.

XX 11-MAY-2000.

XX 18-OCT-1999; 99WO-EP07874.

XX 30-OCT-1998; 98IT-MI02330.

XX (MENA) MENARINI RICERCHE SPA.

XX Parente D, Di Massimo AM, De Santis R;

XX WPI; 2000-365410/31.

XX N-PSDB; AAD00392.

PT Composition containing one or more DNA molecules encoding fragments of
 PT a Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in
 PT anti-tumor therapy -

PS Claim 18; Fig 9; 56pp; English.

CC The present sequence is a fusion protein consisting
 CC of human Mucin 1 (MUC-1) fragment fused to UBILacI sequence at the
 CC N-terminus. The UBILacI sequence consists of ubiquitin from MCF7 cell
 CC line and a portion of E. coli beta-galactosidase (LacI). MUC-1 is an
 CC antigenic protein overexpressed in tumor cells. The corresponding
 CC DNA sequence is cloned into a pMR30 expression vector and used in
 CC pharmaceutical composition e.g. vaccine for inducing an antigen-specific
 CC anti-tumour immune response. Composition containing this DNA molecule
 CC is useful in anti-tumour therapy of patients affected with tumours
 CC characterised by high MUC-1 expression.

SQ Sequence 338 AA;

Query Match 100.0%; Score 44; DB 21; Length 338;
 Best Local Similarity 100.0%; Pred. No. 0.48;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMFLQIY 9
 |||||
 DB 230 ISEMFLQIY 238

RESULT 15
 AAR27662
 ID AAR27662 standard; Protein; 348 AA.
 XX
 AC AAR27662;

DT 06-NOV-1992 (first entry)

DE C-terminal region of H23-ETA-T antigen.

XX Transmembrane; human epithelial antigen; Monoclonal antibody H23;

